

Examen Febrero 2009

Problema 1 . - Regresión múltiple.

El fichero PRegEne09.sav, contiene los datos correspondientes a la observación de 21 días de trabajo en una planta química para la oxidación del amonio como una etapa en la producción del ácido nítrico. Las variables observadas son X1 (flujo de aire), X2 (temperatura del agua de refrigeración en °C), X3 (concentración de ácido (%)) e Y (pérdida acumulada, porcentaje del amonio que escapa sin ser absorbido). El objetivo del estudio es encontrar un modelo que explique la pérdida acumulada de amonio en función de las otras tres variables.

- 1. Indica la variable respuesta y los regresores del problema.**

Las variables del problema, ¿presentan datos atípicos?

¿Podemos suponer que nuestra variable respuesta es

Normal? Justifica tus respuestas.

La variable respuesta es la variable “y” y los regresores son x1 x2 y x3.

Responderemos a la cuestión de si presentan datos atípicos, pinchando en Analizar-> Estadísticos Descriptivos -> Explorar, y seleccionaremos las siguientes opciones:

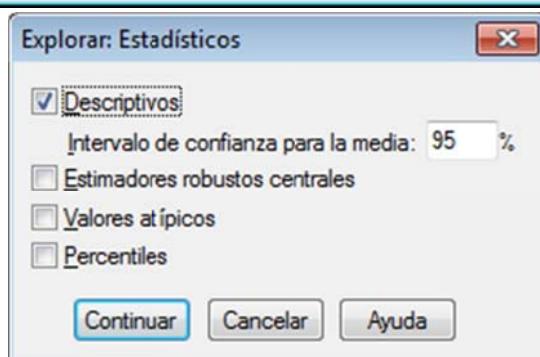
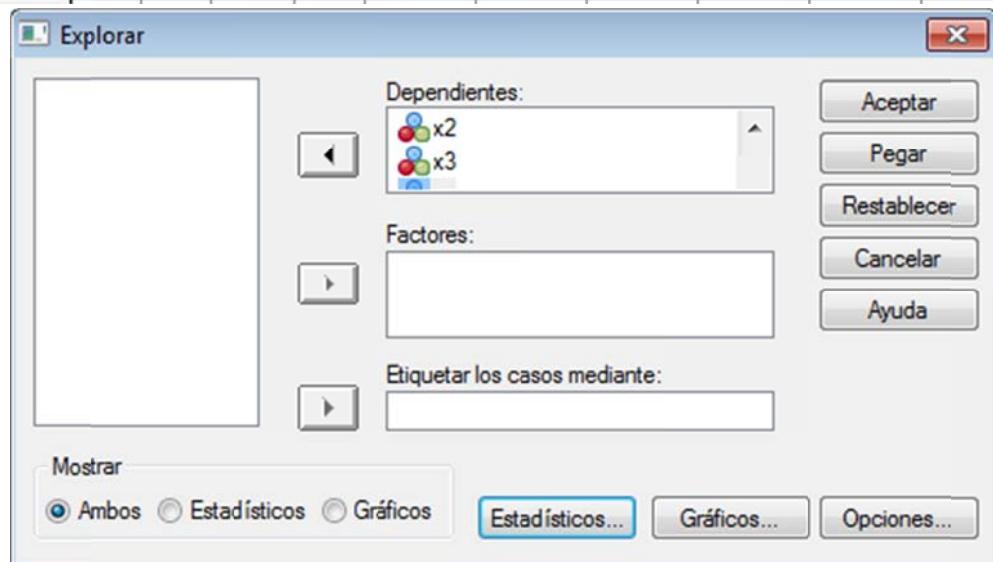
PRegEne09.sav [Conjunto_de_datos1] - Editor de datos SPSS

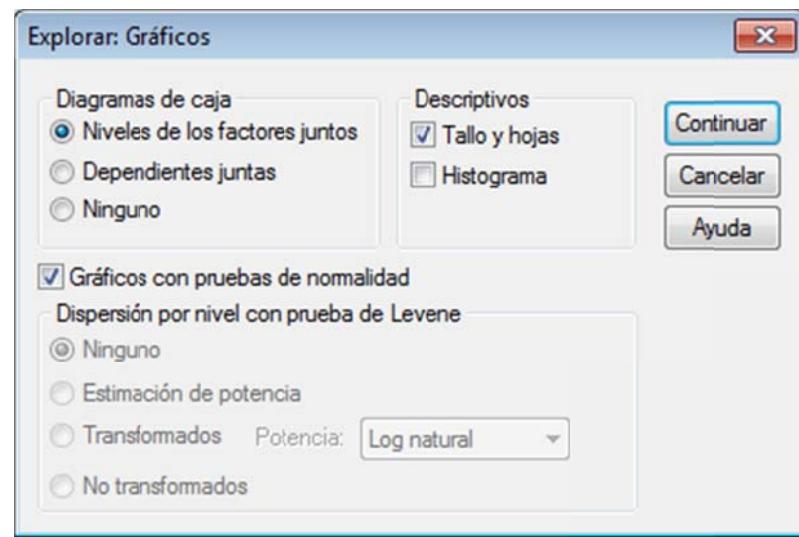
Archivo Edición Ver Datos Transformar Analizar Gráficos Utilidades Ventana ?

1 : x1 80

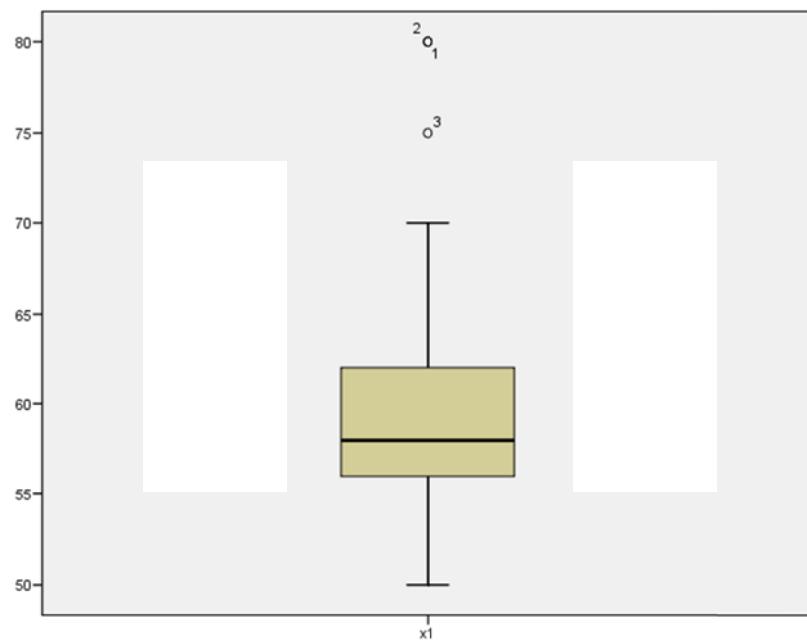
	x1	x2	x3	y
1	80	27	58,9	4,2
2	80	27	58,8	3,7
3	75	25	59,0	3,7
4	62	24	58,7	2,8
5	62	22	58,7	1,8
6	62	23	58,7	1,8
7	62	24	59,3	1,9
8	62	24	59,3	2,0
9	58	23	58,7	1,5
10	58	18	58,0	1,4
11	58	18	58,9	1,4
12	58	17	58,8	1,3
13	58	18	58,2	1,1
14	58	19	59,3	1,2
15	50	18	58,9	,8
16	50	18	58,6	,7
17	50	19	57,2	,8
18	50	19	57,9	,8
19	50	20	58,0	,9
20	56	20	58,2	1,5
21	70	20	59,1	1,5

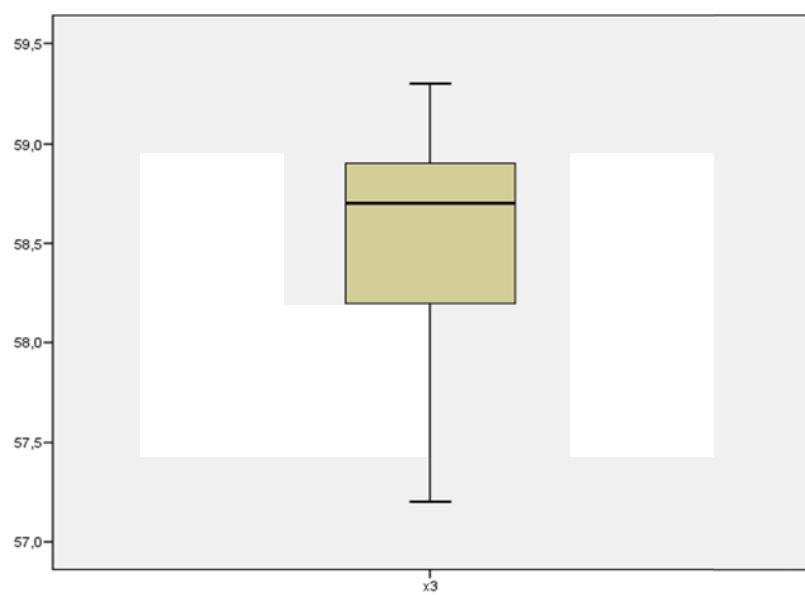
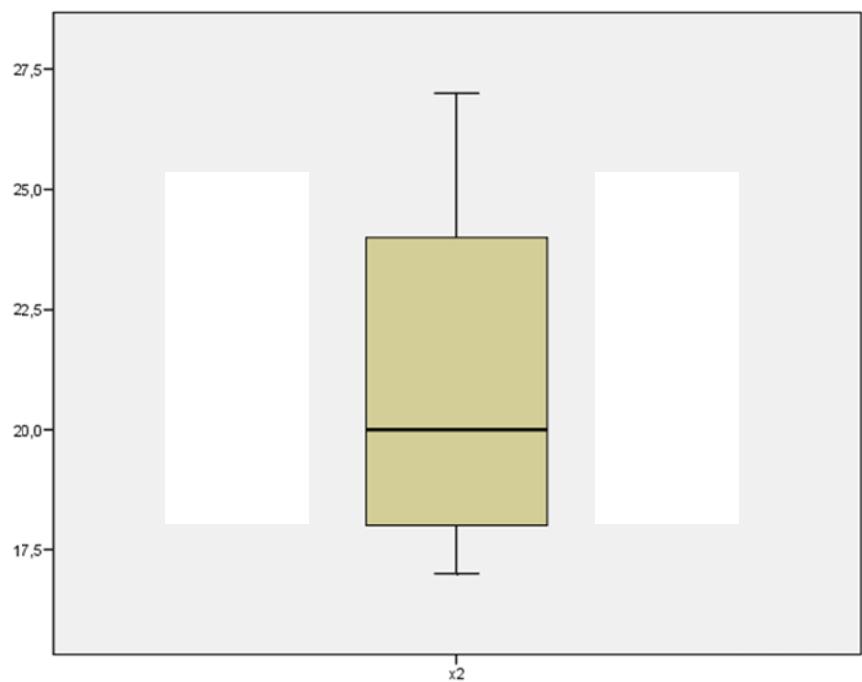
Analizar > Explorar...

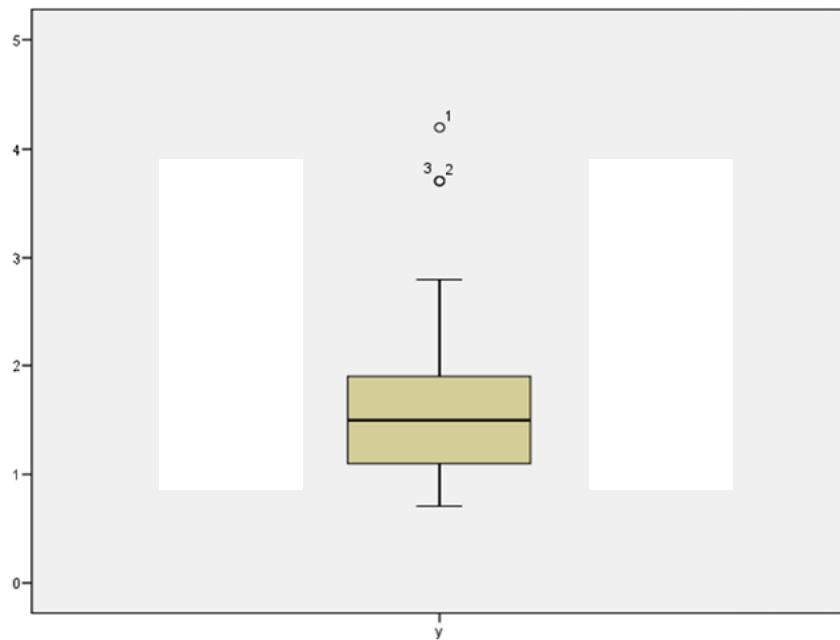




El programa nos devolverá las siguientes graficas:







De las cuales podemos observar 3 valores atípicos (observaciones 1, 2 y 3) que podrían ser valores influyentes del modelo de regresión que queremos proponer. Como no es seguro que sean influyentes, no se eliminarán.

Con respecto a las pruebas de normalidad, utilizaremos la herramienta K-S (Kolmogorov-Smirnov) y comprobaremos los p_valores que nos devuelve SPSS.

Pruebas de normalidad

	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Estadístico	gl	Sig.	Estadístico	gl	Sig.
x1	,241	21	,002	,881	21	,007
x2	,207	21	,019	,895	21	,028
x3	,220	21	,009	,907	21	,048
y	,217	21	,011	,827	21	,002

a. Corrección de la significación de Lilliefors

Al realizar las pruebas de normalidad sobre nuestra variable respuesta, no podemos decir que esta sea normal puesto que presenta un p-valor<0.10, por tanto nos plantearemos realizar una transformación

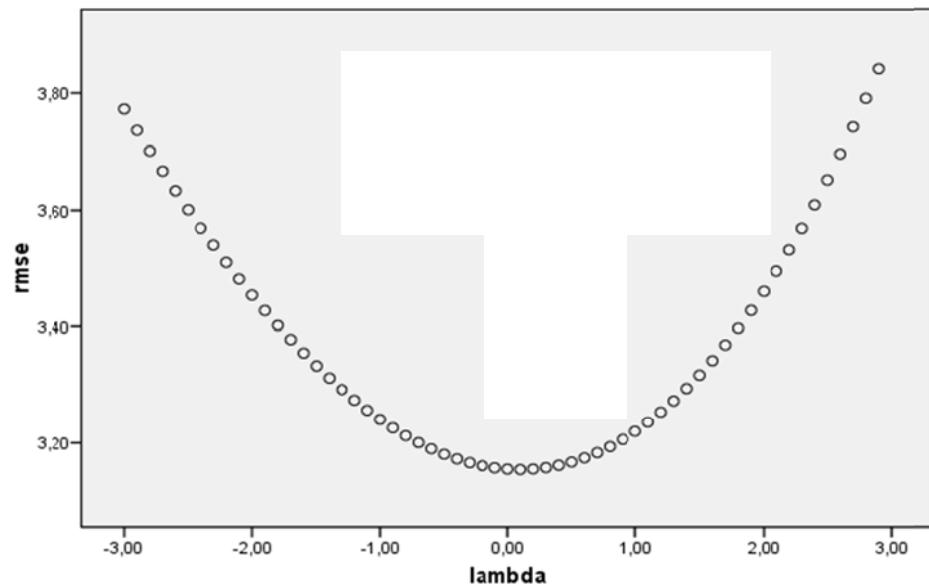
2. En el caso de que la variable respuesta no provenga de una distribución Normal, realiza una transformación adecuada y explica por qué has decidido dicha transformación.

Realizaremos una transformación de box-cox:

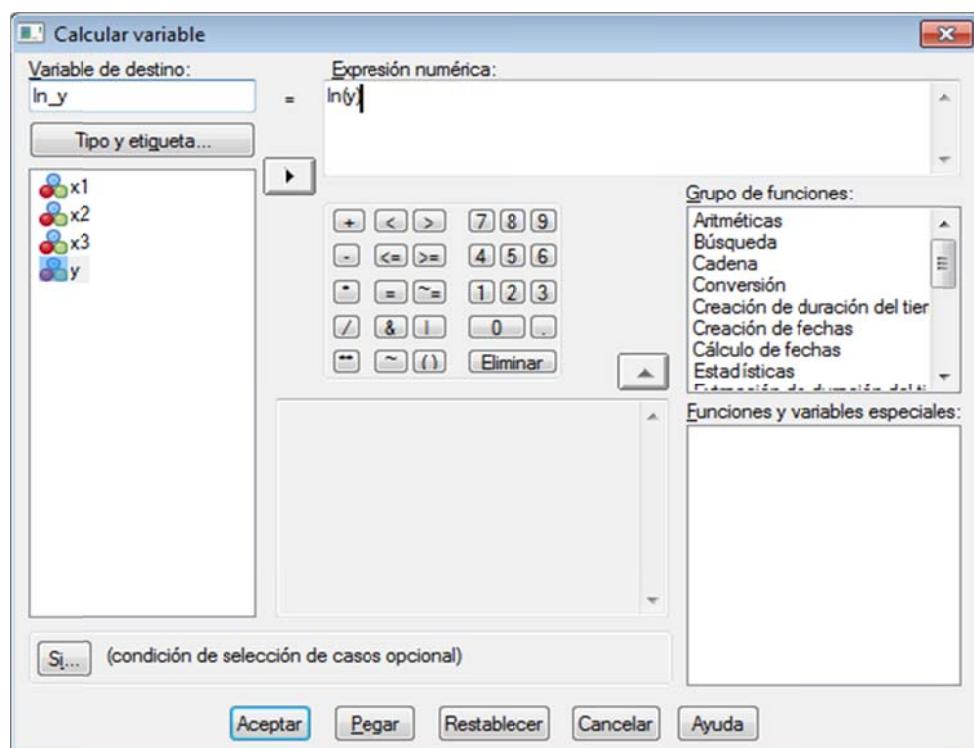
The image shows two windows from the SPSS software. The top window is titled "Abrir sintaxis" (Open syntax) and displays a file selection dialog. It shows a list of files in the "Datos_11" folder, with "box-cox" selected. Below the list are buttons for "Abrir" (Open), "Cancelar" (Cancel), and "Ayuda" (Help). The bottom window is titled "box-cox - Editor de sintaxis SPSS" and shows the syntax code for the Box-Cox transformation. The code includes commands for printing RMS and LAMDA1, computing ANS, looping through NL values, and creating a scatterplot of lambda versus rmse.

```
* PRINT RMS.  
* PRINT LAMDA1.  
  
COMPUTE ANS=MAKE(NL,2,0).  
  
LOOP I = 1 TO NL.  
  COMPUTE ANS(I,1) = LAMDA1(I,1).  
  COMPUTE ANS(I,2)= sqrt(RMS(I,1)).  
END LOOP.  
* PRINT ANS.  
save ANS /outfile=*.  
END MATRIX.  
rename variables col1=lambda col2=rmse.  
GRAPH  
  /SCATTERPLOT(BIVAR)=lambda WITH rmse /MISSING=LISTWISE  
  /TITLE= 'RMSE VERSUS LAMBDA'.
```

RMSE VERSUS LAMBDA



en la que tomaremos logaritmos neperianos de nuestra variable ya que el mínimo de la serie parece que se alcanza entorno al 0.



*PRegEne09.sav [Conjunto_de_datos2] - Editor de datos SPSS

Archivo Edición Ver Datos Transformar Analizar Gráficos

1 : x1 80

	x1	x2	x3	y	ln_y
1	80	27	58,9	4,2	1,44
2	80	27	58,8	3,7	1,31
3	75	25	59,0	3,7	1,31
4	62	24	58,7	2,8	1,03
5	62	22	58,7	1,8	,59
6	62	23	58,7	1,8	,59
7	62	24	59,3	1,9	,64
8	62	24	59,3	2,0	,69
9	58	23	58,7	1,5	,41
10	58	18	58,0	1,4	,34
11	58	18	58,9	1,4	,34
12	58	17	58,8	1,3	,26
13	58	18	58,2	1,1	,10
14	58	19	59,3	1,2	,18
15	50	18	58,9	,8	-,22
16	50	18	58,6	,7	-,36
17	50	19	57,2	,8	-,22
18	50	19	57,9	,8	-,22
19	50	20	58,0	,9	-,11
20	56	20	58,2	1,5	,41
21	70	20	59,1	1,5	,41
22					

Prueba de Kolmogorov-Smirnov para una muestra

N	ln_y
Parámetros normales ^{a,b}	
Media	,4233
Desviación típica	,52353
Diferencias más extremas	
Absoluta	,133
Positiva	,133
Negativa	-,097
Z de Kolmogorov-Smirnov	,608
Sig. asintót. (bilateral)	,854

a. La distribución de contraste es la Normal.

b. Se han calculado a partir de los datos.

Al realizar las pruebas de normalidad sobre nuestra variable ya transformada, podemos afirmar que se comporta como una normal puesto que su p-valor es 0.854 muy superior al p-valor de 0.10.

3. Calcula la matriz de correlaciones de las variables del problema. ¿Qué regresores del modelo presentan una más estrecha relación lineal entre sí? (indica el valor que toma el coeficiente de correlación de Pearson). ¿Cuál es la primera variable que debería entrar en el modelo? (indica el valor que toma el coeficiente de correlación de Pearson).

Resolveremos la pregunta pinchando en Analizar-> Correlaciones -> Bivariadas



Y analizaremos los valores que nos devuelve el programa :

		Correlaciones			
		x1	x2	x3	ln_y
x1	Correlación de Pearson	1	.782**	.500*	.920**
	Sig. (bilateral)		,000	,021	,000
	N	21	21	21	21
x2	Correlación de Pearson	,782**	1	,391	,888**
	Sig. (bilateral)	,000		,080	,000
	N	21	21	21	21
x3	Correlación de Pearson	,500*	,391	1	,482*
	Sig. (bilateral)	,021	,080		,027
	N	21	21	21	21
ln_y	Correlación de Pearson	,920**	,888**	,482*	1
	Sig. (bilateral)	,000	,000	,027	
	N	21	21	21	21

**. La correlación es significativa al nivel 0,01 (bilateral).

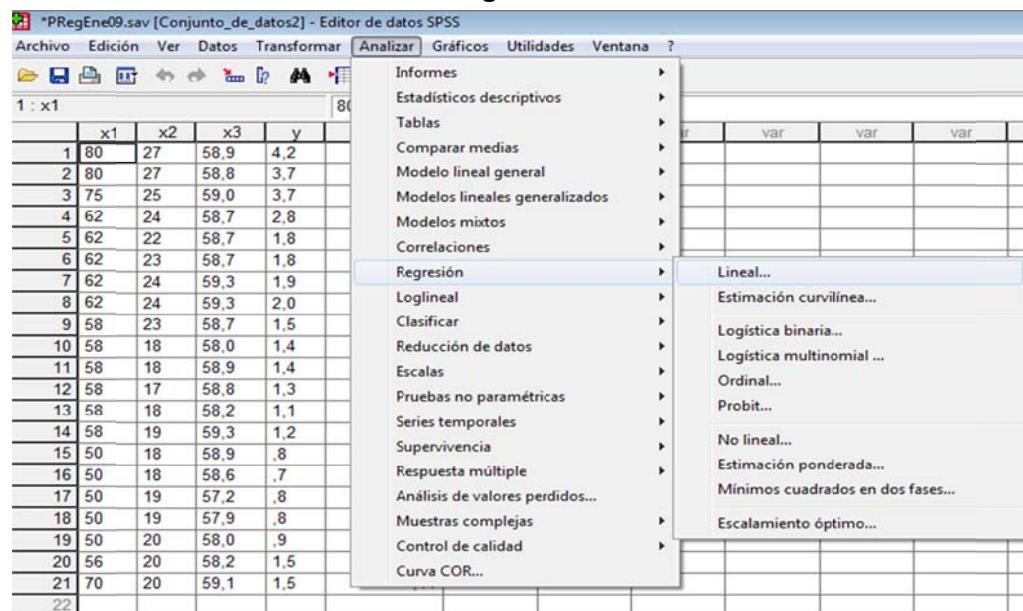
*. La correlación es significante al nivel 0,05 (bilateral).

Se observa que el regresor X1 presenta una alta relación lineal con la variable respuesta, con un coeficiente de correlación de Pearson de 0.92, por lo que deducimos que la primera variable o regresor que debería entrar en el modelo

debería ser el regresor X1. Además, X1 y X2 son los regresores que presentan mayor correlación entre sí, con un coeficiente de 0.782.

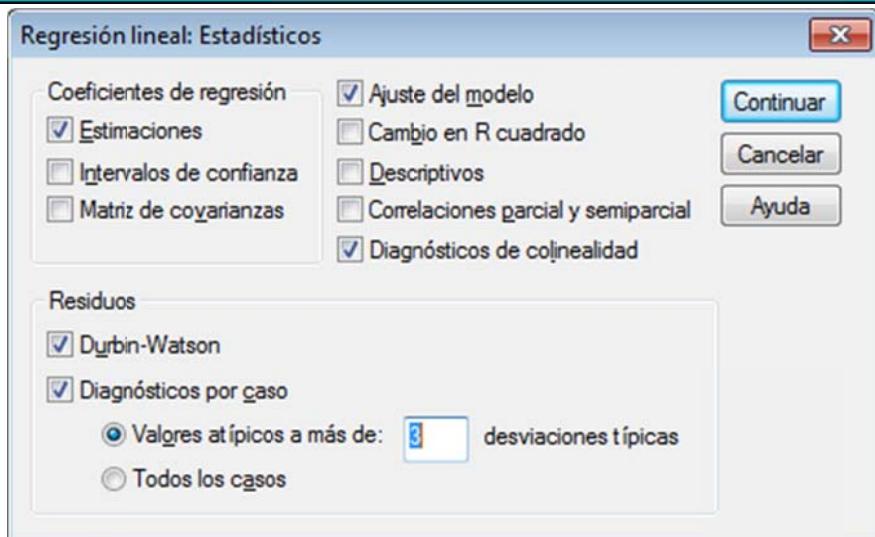
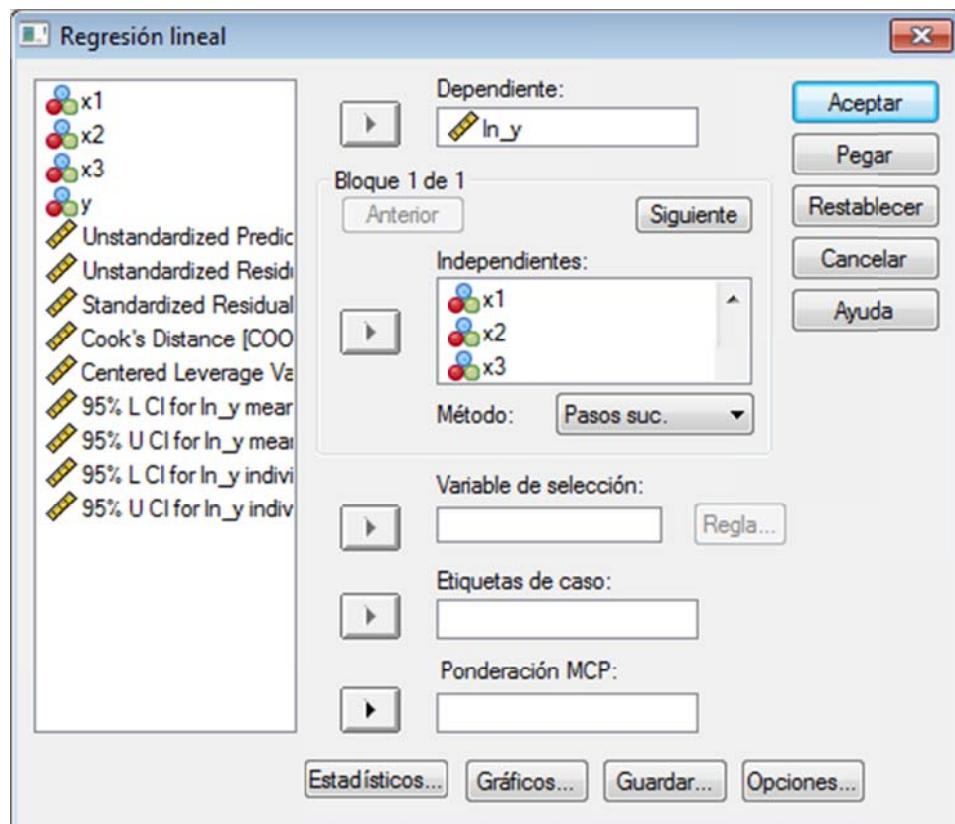
4. Realiza la selección del modelo mediante regresión por pasos, hacia delante y hacia atrás. Indica, para cada uno de los tres métodos, el modelo teórico resultante. Estudia si los modelos obtenidos son reducibles (simplificables) y si presentan multicolinealidad.

Para estudiar los distintos métodos de análisis de nuestro modelo, deberemos meternos en el menú: Analizar -> Regresión -> Lineal.



Dónde iremos variando el método de estimación, usando como variable dependiente el logaritmo neperiano de "y".

- Método por pasos:



Regresión lineal: Guardar nuevas variables

Valores pronosticados	Residuos	<input type="button" value="Continuar"/> <input type="button" value="Cancelar"/> <input type="button" value="Ayuda"/>
<input checked="" type="checkbox"/> No tipificados <input type="checkbox"/> Tipificados <input type="checkbox"/> Corregidos <input type="checkbox"/> E.T. del pronóstico promedio	<input checked="" type="checkbox"/> No tipificados <input checked="" type="checkbox"/> Tipificados <input type="checkbox"/> Estudentizados <input type="checkbox"/> Eliminados <input type="checkbox"/> Eliminados estudiantizados	
Distancias	Estadísticos de influencia	
<input type="checkbox"/> Mahalanobis <input checked="" type="checkbox"/> De Cook <input checked="" type="checkbox"/> Valores de influencia	<input type="checkbox"/> DfBetas <input type="checkbox"/> DfBetas tipificadas <input type="checkbox"/> DfAjuste <input type="checkbox"/> DfAjuste tipificado <input type="checkbox"/> Razón entre covarianzas	
Intervalos de pronóstico		
<input checked="" type="checkbox"/> Media <input checked="" type="checkbox"/> Individuos Intervalo de confianza: <input type="text" value="95"/>		
Estadísticos de los coeficientes		
<input type="checkbox"/> Crear coeficientes de los estadísticos <input checked="" type="radio"/> Crear un nuevo conjunto de datos Nombre de conjunto de datos: <input type="text"/> <input type="radio"/> Escribir un nuevo archivo de datos <input type="button" value="Archivo..."/>		
Exportar información del modelo a un archivo XML		
<input type="text"/> <input type="button" value="Examinar"/>		
<input checked="" type="checkbox"/> Incluir la matriz de covarianzas		

Regresión lineal: Opciones

Criterios del método por pasos	<input type="button" value="Continuar"/> <input type="button" value="Cancelar"/> <input type="button" value="Ayuda"/>
<input checked="" type="radio"/> Usar probabilidad de F Entrada: <input type="text" value=".05"/> Salida: <input type="text" value=".10"/> <input type="radio"/> Usar valor de F Entrada: <input type="text" value="3,84"/> Salida: <input type="text" value="2,71"/>	
<input checked="" type="checkbox"/> Incluir constante en la ecuación	
Valores perdidos	
<input checked="" type="radio"/> Excluir casos según lista <input type="radio"/> Excluir casos según pareja <input type="radio"/> Reemplazar por la media	

Variables introducidas/eliminadas

Modelo	Variables introducidas	Variables eliminadas	Método
1	x1	.	Por pasos (criterio: Prob. de F para entrar <= ,050, Prob. de F para salir >= ,100).
2	x2	.	Por pasos (criterio: Prob. de F para entrar <= ,050, Prob. de F para salir >= ,100).

a. Variable dependiente: ln_y

Resumen del modelo

Modelo	R	R cuadrado	R cuadrado corregida	Error típ. de la estimación	Durbin-Watson
1	,920 ^a	,846	,837	,21108	
2	,950 ^b	,903	,892	,17219	1,425

a. Variables predictoras: (Constante), x1

b. Variables predictoras: (Constante), x1, x2

c. Variable dependiente: ln_y

- Obtenemos una bondad del ajuste del 90,3%, El valor del estadístico de Durbin Watson indica nos hace pensar en independencia de los residuos .

Coefficientes^a

Modelo	Coeficientes no estandarizados		Beta	t	Sig.	Estadísticos de colinealidad	
	B	Error típ.				Tolerancia	FIV
1 (Constante)	-2,750	,314	,920	-8,744	,000	1,000	1,000
x1	,053	,005		10,200	,000		
2 (Constante)	-3,055	,273	,620	-11,182	,000	,389	2,573
x1	,035	,007		5,256	,000		
x2	,063	,020		,383	,248		

a. Variable dependiente: ln_v

- El FIV(factor de varianza inflada) es inferior a 7 en todos los casos, así que nuestro modelo no presentara problemas de multicolinealidad .

Variables excluidas^a

Modelo	Beta dentro	t	Sig.	Correlación parcial	Estadísticos de colinealidad			
					Tolerancia	FIV	Tolerancia mínima	
1	x2	,383 ^b	3,248	,004	,608	,389	2,573	,389
	x3	,029 ^b	,274	,787	,064	,750	1,334	,750
2	x3	,029 ^b	,337	,741	,081	,750	1,334	,344

a. Variables predictoras en el modelo: (Constante), x1

b. Variables predictoras en el modelo: (Constante), x1, x2

c. Variable dependiente: ln_y

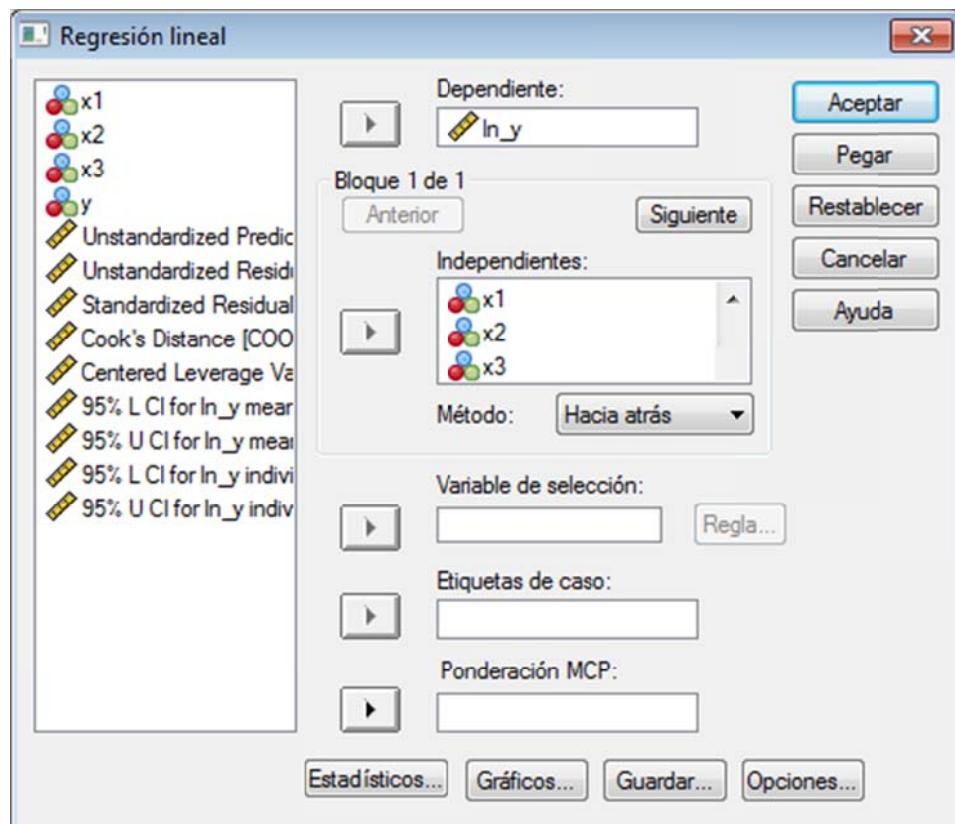
Estadísticos sobre los residuos^b

	Mínimo	Máximo	Media	Desviación tip.	N
Valor pronosticado	-,1423	1,4909	,4233	,49739	21
Valor pronosticado tip.	-1,137	2,146	,000	1,000	21
Error típico del valor pronosticado	,039	,091	,064	,014	21
Valor pronosticado corregido	-,1320	1,5621	,4286	,50683	21
Residuo bruto	-,28720	,38632	,00000	,16336	21
Residuo tip.	-1,868	2,127	,000	,949	21
Residuo estud.	-1,980	2,279	-,014	1,039	21
Residuo eliminado	-,39670	,42028	-,00531	,19653	21
Residuo eliminado estud.	-2,148	2,825	-,003	1,102	21
Dist. de Mahalanobis	,089	4,659	1,905	1,329	21
Distancia de Cook	,000	,488	,071	,117	21
Valor de influencia centrado	,004	,233	,095	,066	21

a. Variable dependiente: ln_y

- Los residuos tipificados no se salen del intervalo(-3,3) así que no podemos considerar que existan residuos atípicos.
- El máximo valor para la distancia de Cook es de 0.488 y el máximo valor de influencia centrado es 0,233 por tanto podemos decir que no existen observaciones influyentes.

- Método hacia atrás:



Variables introducidas/eliminadas^b

Modelo	Variables introducidas	Variables eliminadas	Método
1	x3, x2, x1	.	Introducir
2	.	x3	Hacia atrás (criterio: Prob. de F para salir $\geq .100$).

a. Todas las variables solicitadas introducidas

b. Variable dependiente: ln_y

Resumen del modelo^c

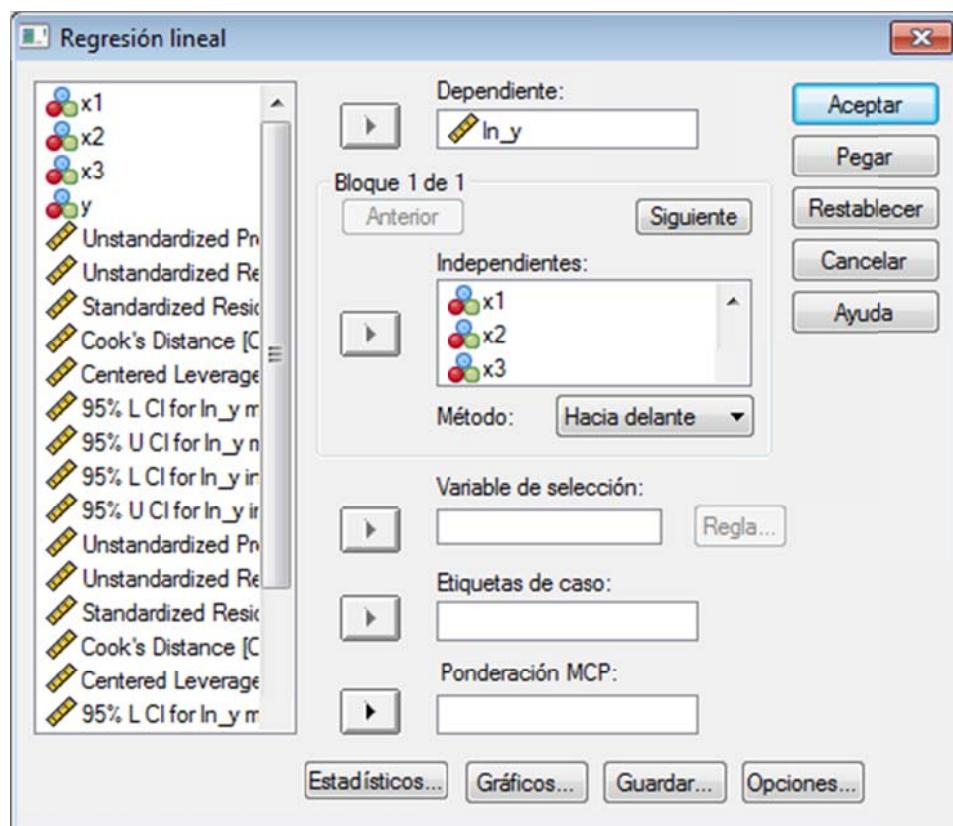
Modelo	R	R cuadrado	R cuadrado corregida	Error típ. de la estimación	Durbin-Watson
1	,950 ^a	,903	,888	,17680	
2	,950 ^b	,903	,892	,17219	1,425

a. Variables predictoras: (Constante), x3, x2, x1

b. Variables predictoras: (Constante), x2, x1

c. Variable dependiente: ln_y

- Método hacia delante:



Modelo	Variables introducidas	Variables eliminadas	Método
1	x1		Hacia adelante (criterio: Prob. de F para entrar <= ,050)
2	x2		Hacia adelante (criterio: Prob. de F para entrar <= ,050)

a Variable dependiente: ln_y

Para todos los métodos estudiados, los datos sobre los modelos que nos arroja SPSS son idénticos, así que a partir de ahora nos remitiremos a usar un único modelo, por ejemplo el de pasos sucesivos.

Concluimos que para todos los métodos de estudio posibles los resultados que nos arroja SPSS no presentan multicolinealidad, puesto que el FIV es inferior a 7 en todos los casos posibles.

Los modelos obtenidos nos son reducibles puesto que los p_valores para los regresores que han entrado dentro del modelo, indican que son significativos.

5. ¿Qué modelo de regresión propondrías y por qué?

Determina el modelo ajustado y comenta el valor de R².

Como hemos podido observar realizando el método de regresión por pasos, hacia adelante o hacia atrás, y dicho anteriormente, el SPSS nos devuelve unas modelos iguales utilizando estos métodos.

Propondríamos un modelo con los regresores X1 y X2 puesto que son los que presentan mayor relación con la variable respuesta, obteniendo un coeficiente de R² de 0.903 (bondad del ajuste).

El modelo ajustado sería el siguiente → $\ln(y) = -3.055 + 0.035X_1 + 0.063X_2 + \varepsilon$

6. Para el modelo resultante del apartado anterior, guarda los residuos y los valores ajustados y estudia si se verifican las hipótesis del modelo de regresión múltiple, comentando los procesos utilizados.

- Hipótesis de normalidad: realizaremos un contraste de normalidad para los residuos, como observamos el p_valor indica que validamos la hipótesis de normalidad puesto que este es mayor de 0.10.

*PRegEne09.sav [Conjunto_de_datos2] - Editor de datos SPSS

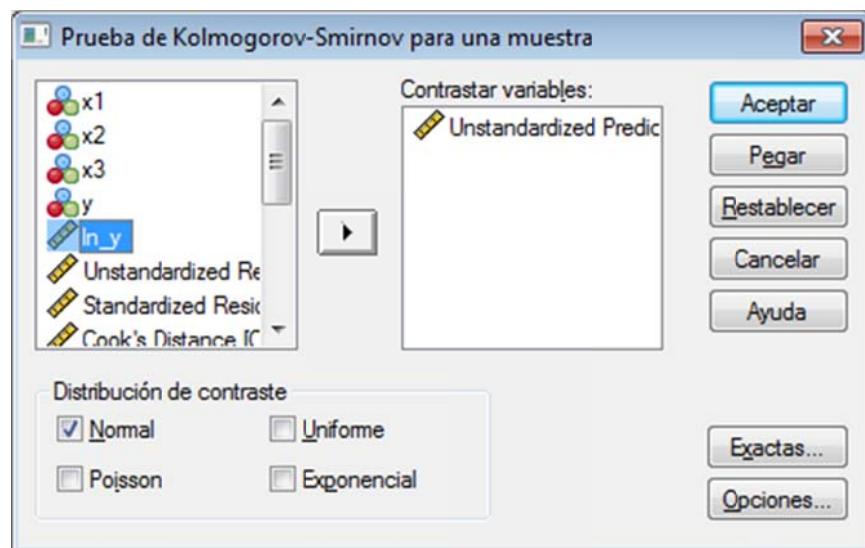
Archivo Edición Ver Datos Transformar Analizar Gráficos Utilidades Ventana ?

	y	ln_y	PRE
1	4,2	1,44	1
2	3,7	1,31	1
3	3,7	1,31	1
4	2,8	1,03	
5	1,8	,59	
6	1,8	,59	
7	1,9	,64	
8	2,0	,69	
9	1,5	,41	
10	1,4	,34	
11	1,4	,34	
12	1,3	,26	
13	1,1	,10	
14	1,2	,18	
15	,8	-,22	
16	,7	-,36	
17	,8	-,22	
18	,8	-,22	
19	,9	-,11	
20	1,5	,41	
21	1,5	,41	
22			

Informes
Estadísticos descriptivos
Tablas
Comparar medias
Modelo lineal general
Modelos lineales generalizados
Modelos mixtos
Correlaciones
Regresión
Loglineal
Clasificar
Reducción de datos
Escalas
Pruebas no paramétricas
Series temporales
Supervivencia
Respuesta múltiple
Análisis de valores perdidos...
Muestras complejas
Control de calidad
Curva COR...

	COO_1	LEV_1
436	,01901	,23
046	,20320	,23
465	,04218	,12
737	,25485	,08
857	,00172	,00
000	,00015	,02
457	,00087	,08
332	,00169	,08
643	,00567	,08
571	,07224	,08
571	,07224	,08

Chi-cuadrado...
Binomial...
Rachas...
K-S de 1 muestra...
2 muestras independientes...
K muestras independientes...
2 muestras relacionadas...
K muestras relacionadas...



¡OJO! AQUÍ HAY UN ERROR PORQUE SE HA REALIZADO LA PRUEBA DE NORMALIDAD A LOS VALORES AJUSTADOS EN LUGAR DE LOS RESIDUOS.

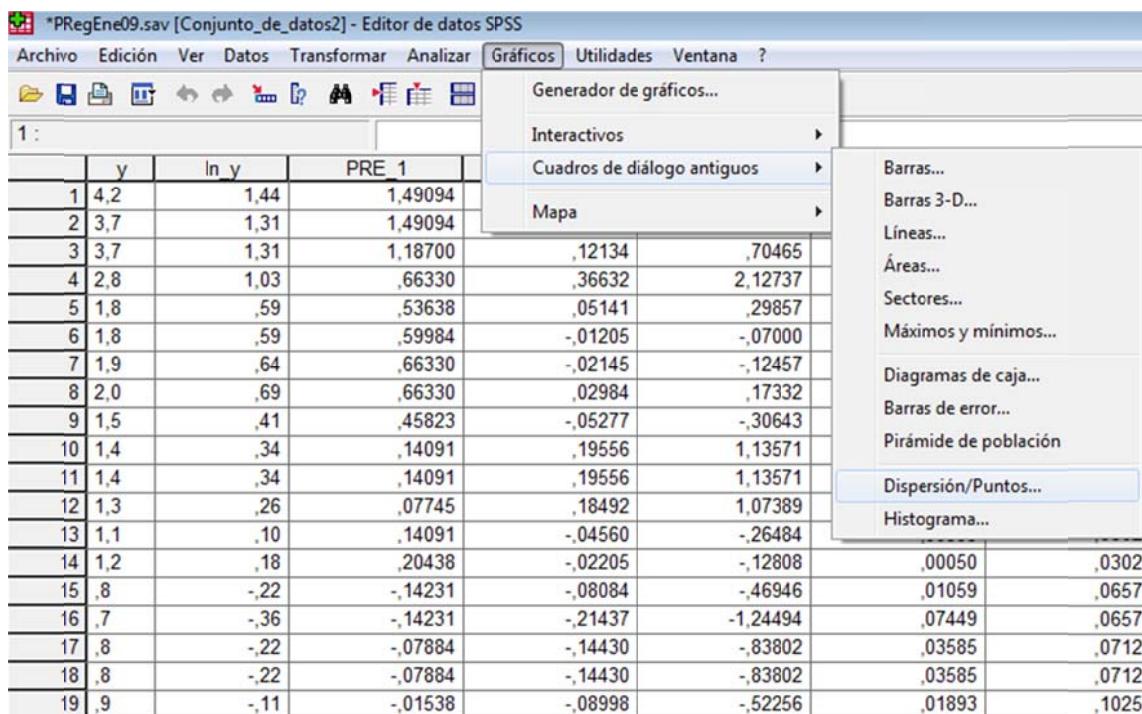
Prueba de Kolmogorov-Smirnov para una muestra

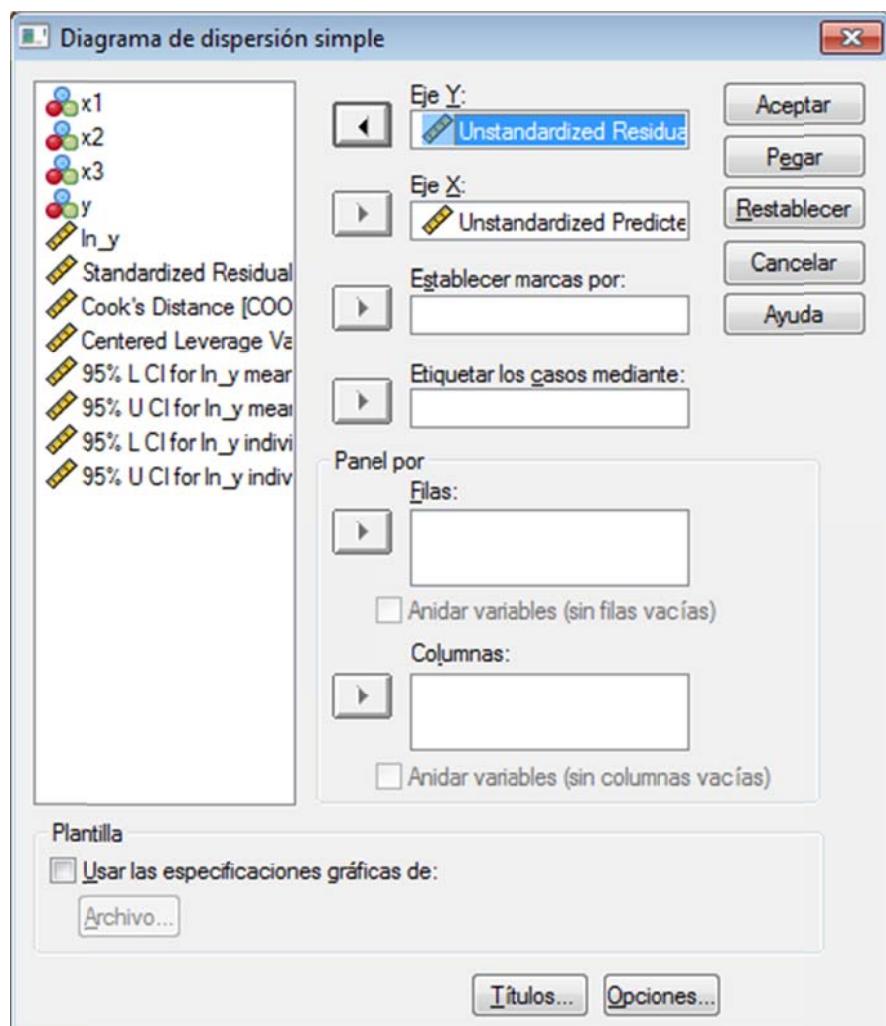
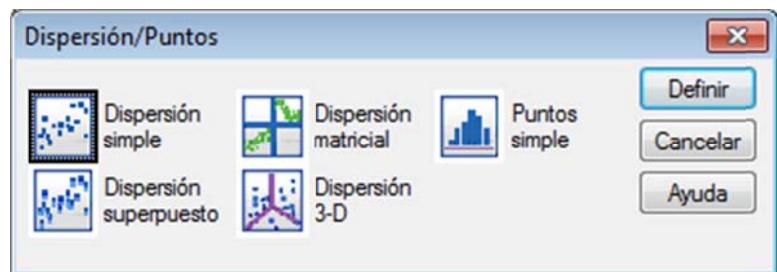
		Unstandardized Predicted Value
N		21
Parámetros normales ^{a,b}	Media	,4233245
	Desviación típica	,49739225
Diferencias más extremas	Absoluta	,194
	Positiva	,194
	Negativa	-,128
Z de Kolmogorov-Smirnov		,889
Sig. asintót. (bilateral)		,409

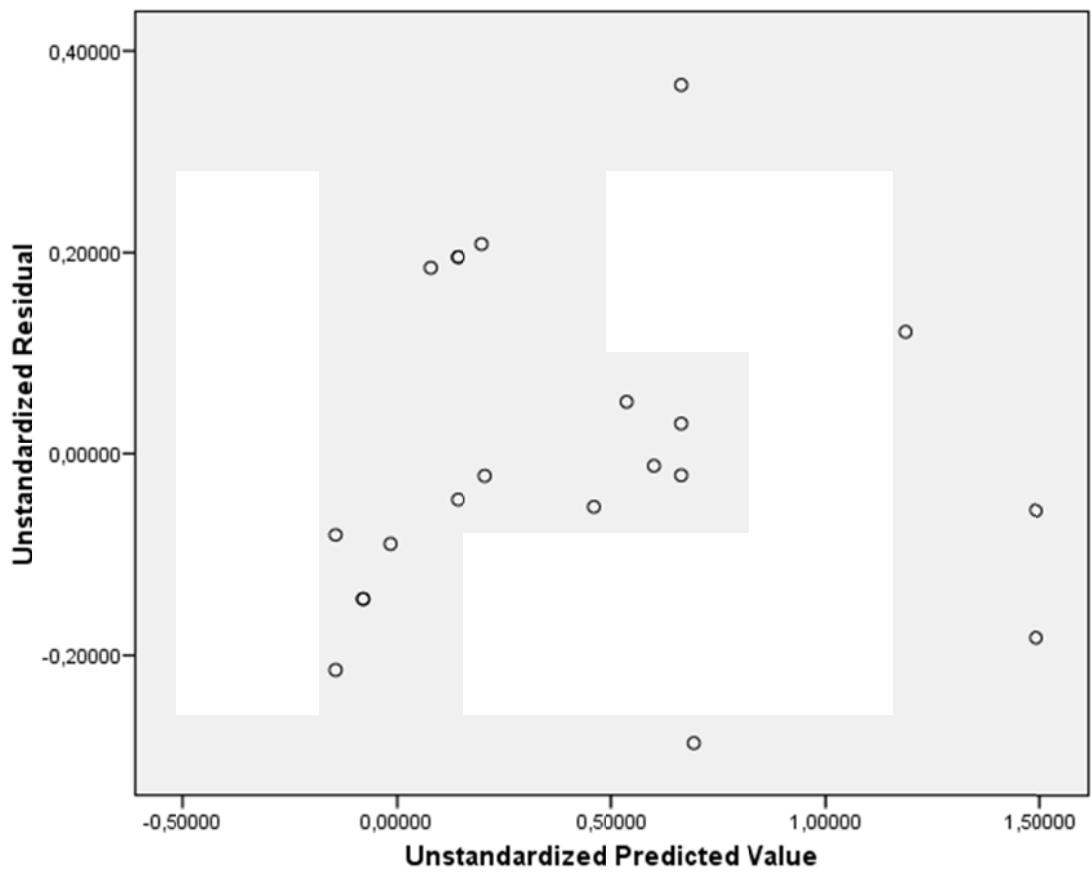
a. La distribución de contraste es la Normal.

b. Se han calculado a partir de los datos.

- Hipótesis de homocedasticidad: para validar esta hipótesis representaremos los residuos en función de los valores ajustados en un gráfico de dispersión, y observamos que presenta una disposición de datos de forma aleatoria por tanto podemos validar esta hipótesis.







- Hipótesis de independencia: para comprobar esta hipótesis Realizaremos un correlograma o autocorrelograma, en el cual podemos observar que no existe ninguna correlación significativa.

*PRegEne09.sav [Conjunto_de_datos2] - Editor de datos SPSS

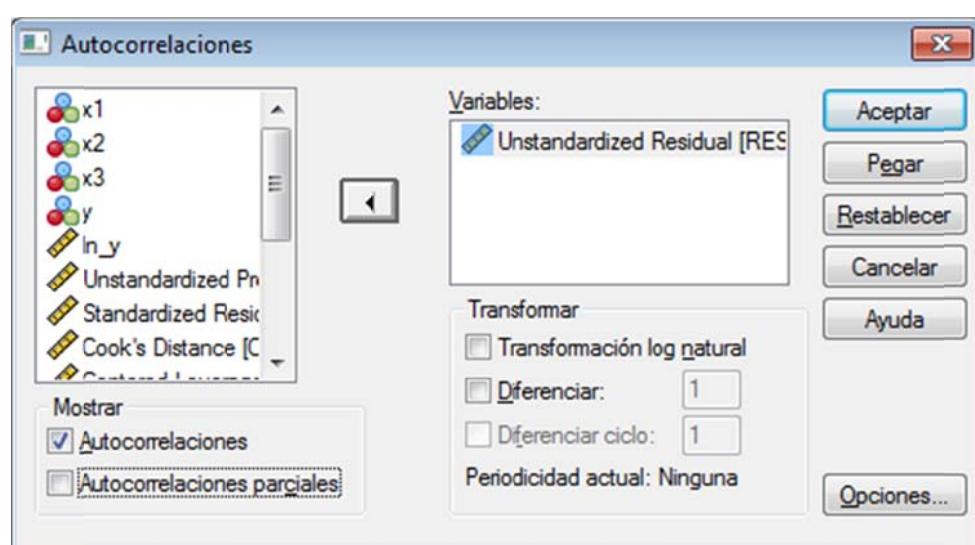
Archivo Edición Ver Datos Transformar Analizar Gráficos Utilidades Ventana ?

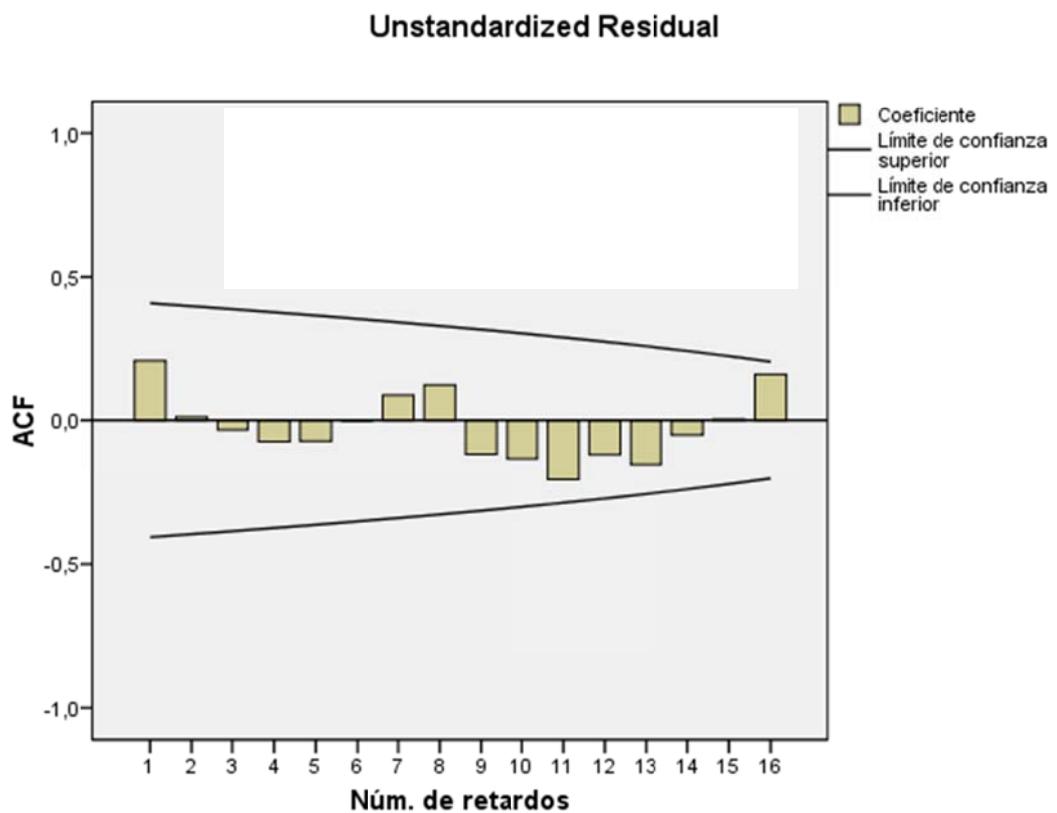
1 :

	y	ln y	PRE
1	4,2	1,44	1
2	3,7	1,31	1
3	3,7	1,31	1
4	2,8	1,03	
5	1,8	.59	
6	1,8	.59	
7	1,9	.64	
8	2,0	.69	
9	1,5	.41	
10	1,4	.34	
11	1,4	.34	
12	1,3	.26	
13	1,1	.10	
14	1,2	.18	
15	.8	-.22	
16	.7	-.36	
17	.8	-.22	
18	.8	-.22	
19	.9	-.11	
20	1,5	.41	
21	1,5	.41	
22			
23			
24			
25			
26			
27			

Analizar menu open:

- Informes
- Estadísticos descriptivos
- Tablas
- Comparar medias
- Modelo lineal general
- Modelos lineales generalizados
- Modelos mixtos
- Correlaciones
- Regresión
- Loglineal
- Clasificar
- Reducción de datos
- Escalas
- Pruebas no paramétricas
- Series temporales
- Supervivencia
- Respuesta múltiple
- Análisis de valores perdidos...
- Muestras complejas
- Control de calidad
- Curva COR...
- Crear modelos...
- Aplicar modelos...
- Suavizado exponencial...
- Autorregresión...
- ARIMA...
- Descomposición estacional...
- Análisis espectral...
- Gráficos de secuencia...
- Autocorrelaciones...
- Correlaciones cruzadas...





7. ¿Existen observaciones influyentes? Comenta los criterios utilizados.

El máximo valor para la distancia de Cook es 0.488 y el máximo valor de influencia centrado es 0.233 por tanto podemos afirmar que no existen observaciones influyentes.

Los criterios que hemos utilizado han sido la distancia de Cook y el Valor de influencia centrado. En los apuntes se puede consultar en qué consisten.

iFALTARÍA COMENTAR QUE LOS RESIDUOS GRANDES PUEDEN INDICAR VALORES INFLUYENTES! RESIDUOS TIPIFICADOS FUERA DEL INTERVALO (-3,3)

8. Proporciona una estimación puntual de la pérdida de amonio cuando $X_1 = 60$, $X_2 = 20$, $X_3 = 60$. ¿Podemos asegurar que la pérdida de amonio será inferior a 1.8? ¿Y en promedio para los días de esas características?.

	x1	x2	x3	y	ln y	Predicción	PRE_1	LMCI_1	UMCI_1	LICI_1	UICI_1
1	80	27	58,9	4,2	1,44	1,45	1,49094	1,29932	1,68256	1,08156	1,90031
2	80	27	58,8	3,7	1,31	1,45	1,49094	1,29932	1,68256	1,08156	1,90031
3	75	25	59,0	3,7	1,31	1,15	1,18700	1,03613	1,33787	.79504	1,57896
4	62	24	58,7	2,8	1,03	.63	.66330	.53370	.79291	.27902	1,04758
5	62	22	58,7	1,8	.59	.50	.53638	.45383	.61893	.16531	.90744
6	62	23	58,7	1,8	.59	.56	.59984	.49924	.70044	.22435	.97533
7	62	24	59,3	1,9	.64	.63	.66330	.53370	.79291	.27902	1,04758
8	62	24	59,3	2,0	.69	.63	.66330	.53370	.79291	.27902	1,04758
9	58	23	58,7	1,5	.41	.42	.45823	.32508	.59138	.07274	.84372
10	58	18	58,0	1,4	.34	.11	.14091	.01157	.27025	.-24328	.52510
11	58	18	58,9	1,4	.34	.11	.14091	.01157	.27025	.-24328	.52510
12	58	17	58,8	1,3	.26	.05	.07745	-.08576	.24066	-.31943	.47433
13	58	18	58,2	1,1	.10	.11	.14091	.01157	.27025	.-24328	.52510
14	58	19	59,3	1,2	.18	.17	.20438	.10344	.30531	-.17120	.57996
15	50	18	58,9	.8	-.22	-.17	-.14231	-.26410	-.02051	-.52402	.23941
16	50	18	58,6	.7	-.36	-.17	-.14231	-.26410	-.02051	-.52402	.23941
17	50	19	57,2	.8	-.22	-.11	-.07884	-.20359	.04590	-.46151	.30382
18	50	19	57,9	.8	-.22	-.11	-.07884	-.20359	.04590	-.46151	.30382
19	50	20	58,0	.9	-.11	-.05	-.01538	-.15559	.12483	-.40336	.37260
20	56	20	58,2	1,5	.41	.17	.19703	.10886	.28521	-.17532	.56939
21	70	20	59,1	1,5	.41	.66	.69267	.50261	.88273	.28402	1,10132
22	60	2031	.33864	.24997	.42732	-.03383	.71112
23											

- Estimación puntual: 0,33864
- Intervalo de confianza respuesta promedio: (0,24997 ; 0,42732).
- Intervalo de predicción para observaciones futuras: (-0,03383; 0,71112)

Estos valores son para la variable respuesta con transformación de logaritmo neperiano, para conocer los valores reales de la respuesta variable, desharemos este cambio tomando la exponencial de los valores anteriores.

- Estimación puntual → 1,403

$$e^{0,33864} = 1,403$$

- Intervalo de confianza respuesta promedio → (1,28 ; 1,53)

$$e^{0,24997} = 1,28$$

$$e^{0,42732} = 1,53$$

- Intervalo de predicción para observaciones futuras → (0,96 ; 2,03)

$$e^{-0,03383} = 0,96$$

$$e^{0,71112} = 2,03$$

No podemos afirmar que la perdida de amonio sea inferior a 1,8 puesto que el intervalo de predicción para observaciones futuras nos indica que podría ser hasta

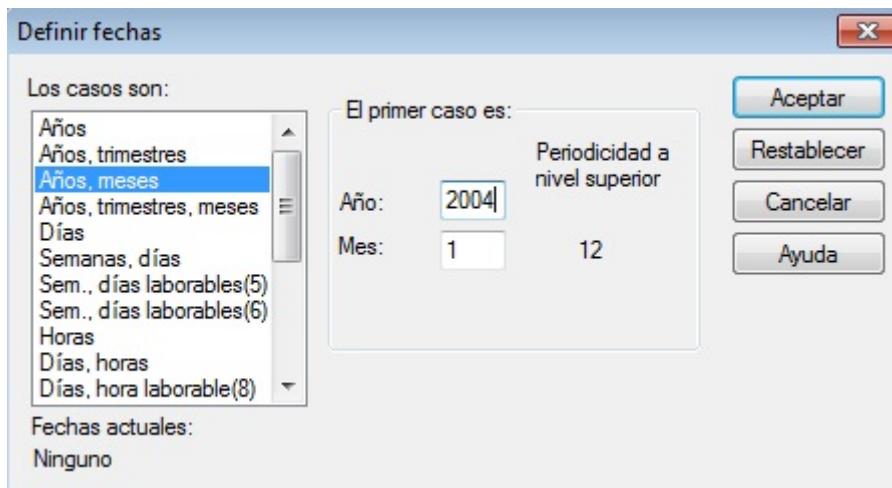
2,03; En cambio para la respuesta promedio si podríamos afirmarlo puesto que el valor máximo que podemos obtener dicha respuesta promedio será de 1,53.

Examen Febrero 2009

Problema 3 . - Análisis clásico de series temporales.

1. Representa los datos de la ocupación hotelera en un grafico temporal. ¿ La serie presenta Estacionalidad?

En primer lugar definiremos las fechas pichando en datos-> definir fechas.



Y elegiremos la opción meses, años, definiendo así, como comienzo de la serie enero de 2004 , observamos que se nos han generado 3 nuevas variables, year_ : ;month_ : ;y Date_ ; y se han guardado en la hoja de datos.

Posteriormente para representar la serie seleccionaremos la opción gráficos de secuencia del menú, series temporales a su vez incluido en analizar(Analizar -> Series Temporales -> Gráficos de secuencia.)

trabajo_hoteles.sav - SPSS Editor de datos

Archivo Edición Ver Datos Transformar Analizar Gráficos Utilidades Ventana ?

1 :

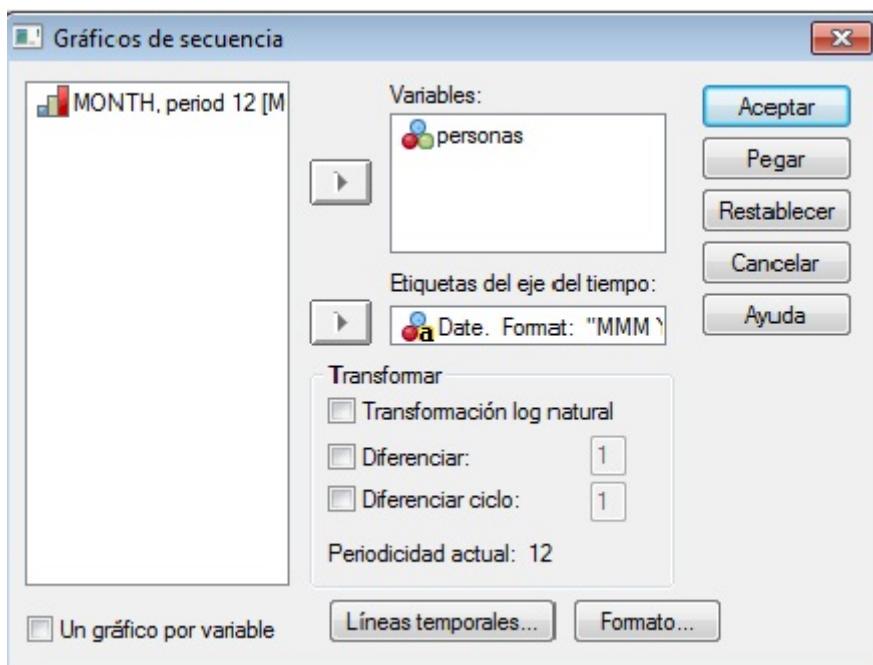
	personas	MONTH	
1	55656	1 JA	
2	71374	2 FE	
3	77550	3 MAR	
4	83860	4 ABR	
5	86920	5 MAY	
6	84441	6 JUN	
7	82357	7 JUL	
8	92653	8 AGO	
9	88372	9 SEPT	
10	77960	10 OCT	
11	67621	11 NOV	
12	57262	12 DICI	
13	55232	1 JA	
14	69725	2 FE	
15	88786	3 MAR	
16	87578	4 ABR	
17	89278	5 MAY	
18	94092	6 JUN	
19	93741	7 JUL	
20	100909	8 AGO	
21	97512	9 SEPT	
22	84857	10 OCT 2005	
23	74779	11 NOV 2005	
24	63457	12 DICI 2005	
25	63835	1 JAN 2006	

Analizar ▾

- Informes ▾
- Estadísticos descriptivos ▾
- Tablas ▾
- Comparar medias ▾
- Modelo lineal general ▾
- Modelos lineales generalizados ▾
- Modelos mixtos ▾
- Correlaciones ▾
- Regresión ▾
- Loglineal ▾
- Clasificar ▾
- Reducción de datos ▾
- Escalas ▾
- Pruebas no paramétricas ▾
- Series temporales ▾
- Supervivencia ▾
- Respuesta múltiple ▾
- Análisis de valores perdidos... ▾
- Muestras complejas ▾
- Control de calidad ▾
- Curva COR... ▾

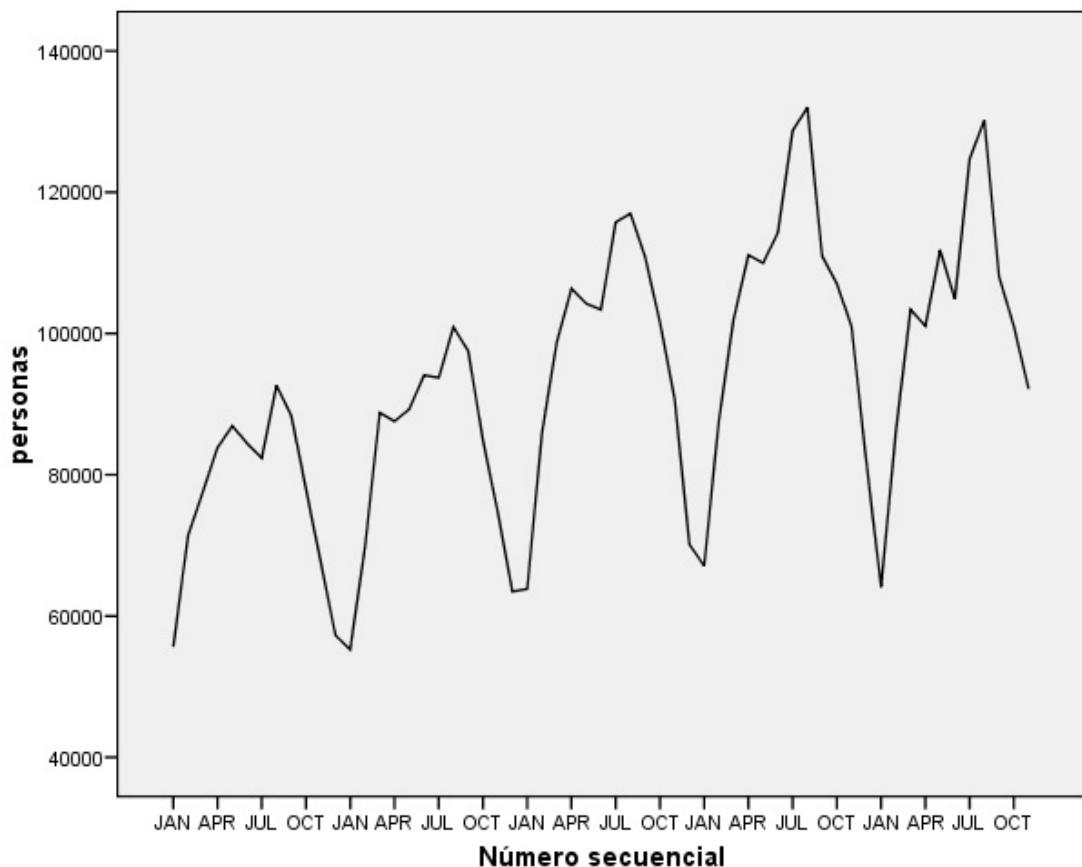
Crear modelos...
Aplicar modelos...
Suavizado exponencial...
Autorregresión...
ARIMA...
Descomposición estacional...
Análisis espectral...
Gráficos de secuencia...
Autocorrelaciones...
Correlaciones cruzadas...

Seleccionaremos como variable personas y como etiquetas del eje del tiempo la variable Date.



Esta opción nos devuelve la siguiente serie de la que podemos deducir, que sí presenta una estacionalidad que se repite cada 12 meses, con una tendencia

aparentemente lineal con una pendiente positiva y que a priori parece la de un modelo multiplicativo.



2. Justifica si se trata de un modelo aditivo o multiplicativo.

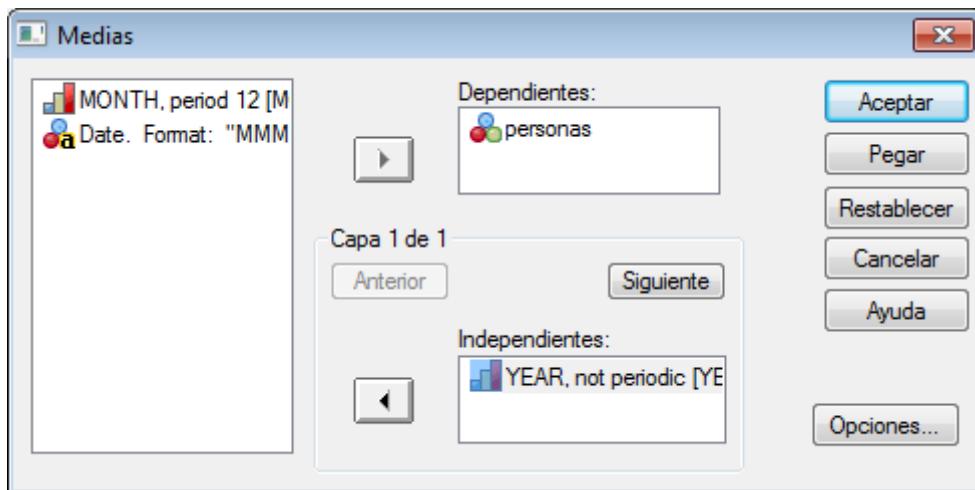
Para confirmarlo, realiza un grafico de desviaciones típicas frente a medias para cada año(represéntalo en tu folio a mano alzada).

Para realizar una comprobación tangible de que realmente es un modelo multiplicativo como inicialmente se observa generamos el grafico comparativo de las desviaciones típicas frente a las medias. Para ello seleccionaremos la opción:

Analizar-> Comparar medias -> Medias.

The screenshot shows the SPSS software interface. The main window displays a data table with two columns: 'personas' and 'MONTH'. The 'MONTH' column contains abbreviations like JA, FE, MAR, APR, MAY, JUN, JUL, AUG, SEPT, OCT, NOV, DEC, JAN, FEB, MAR, APR, MAY, JUN, JUL, AUG, SEPT, OCT, NOV, and SEPT. The data ranges from 1 to 22. The 'Analyze' menu is open, and the 'Comparar medias' (Compare Means) option is selected, with 'Medias...' (Means) highlighted.

Pasaremos la variable dependiente personas y como variable independiente la variable Year. Posteriormente pincharemos en opciones y seleccionaremos únicamente *Media* y *Desviación típica*.



Obteniendo la siguiente tabla de la que seleccionaremos los datos para llevarlos al Editor de datos para poder así trabajar con ellos.

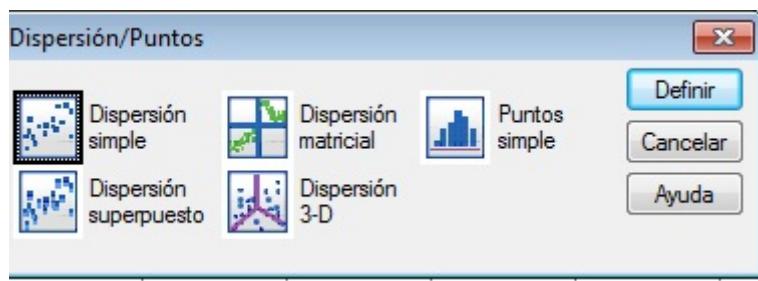
Informe

personas

YEAR, not periodic	Media	N	Desv. típ.
2004	77168,83	12	11945,046
2005	83328,83	12	14341,716
2006	97374,08	12	16872,767
2007	104472,00	12	18528,915
2008	102495,91	11	17989,847
Total	92806,44	59	18985,008

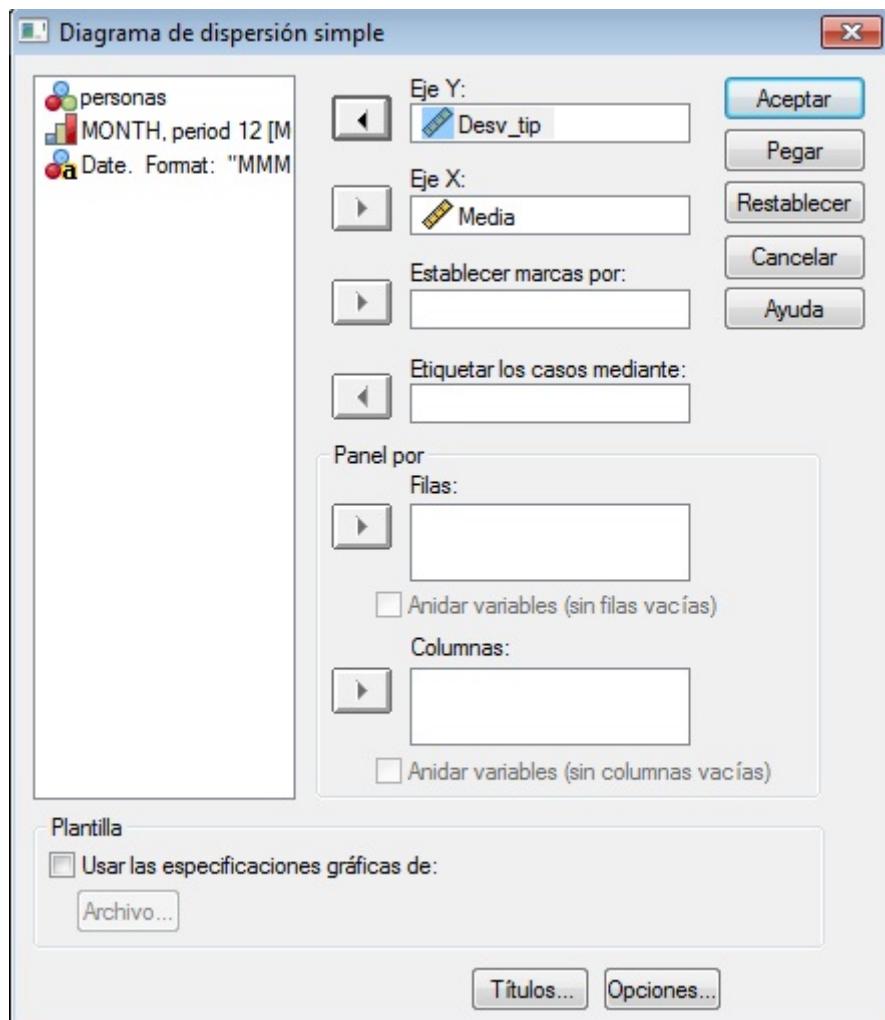
Posteriormente representaremos gráficamente esta información en un gráfico de dispersión de desviaciones típicas frente a medias.

The screenshot shows the SPSS interface with the 'trabajo_hoteles.sav - SPSS Editor de datos' window open. The 'Gráficos' menu is active, and the 'Dispersión/Puntos...' option is highlighted in blue. A dropdown menu on the right lists various chart types, with 'Dispersión/Puntos...' also highlighted in blue.

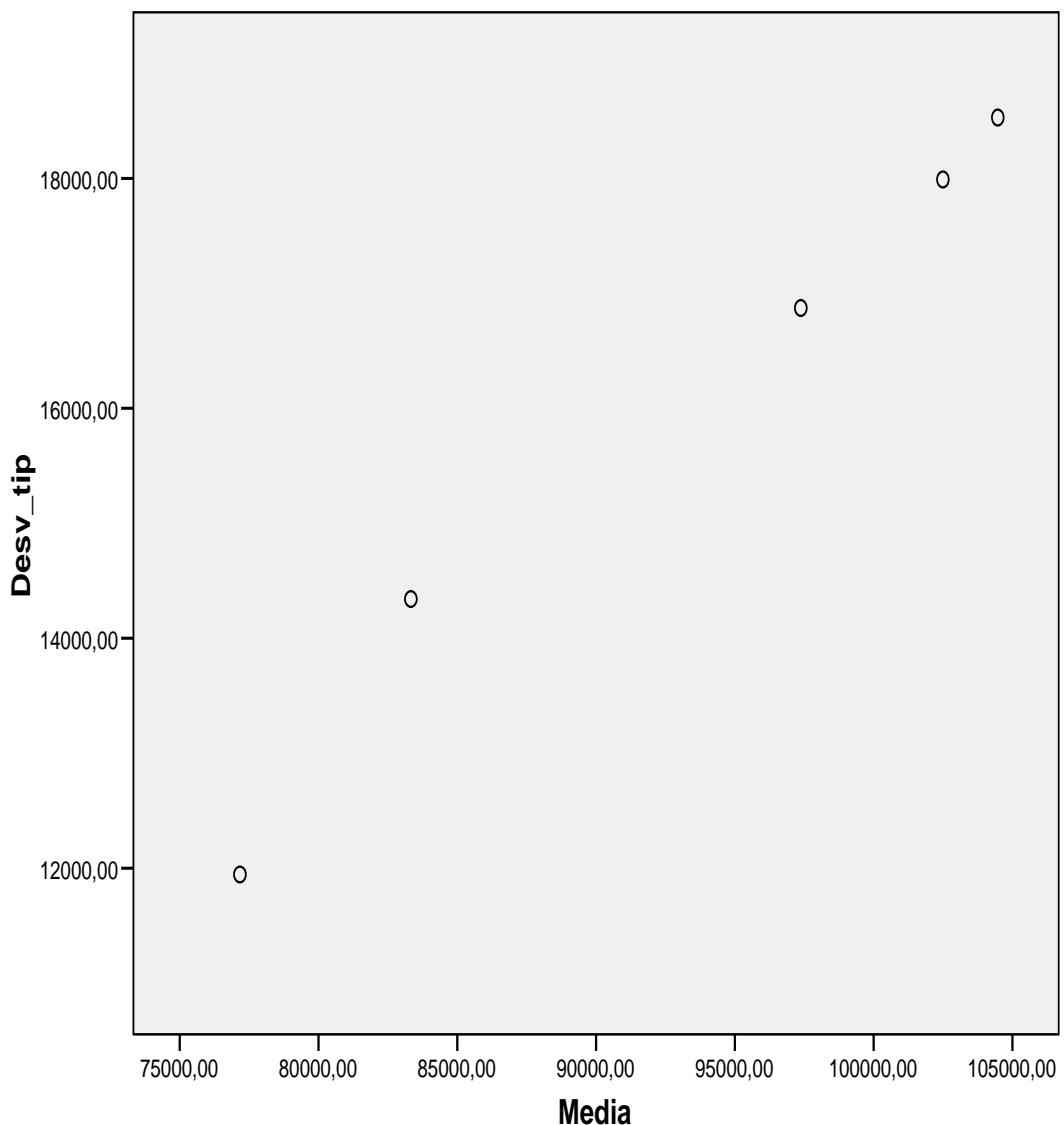


Seleccionaremos:

Gráficos -> Dispersión ->Diagrama de dispersión Simple Pasaremos las variables:
medias en el eje X ,y desviación típica en el eje Y



SPSS nos devuelve la siguiente gráfica:



De la que concluimos que se observa un claro aumento lineal de la desviación típica conforme aumenta la media, y que por tanto, se trata de un esquema multiplicativo, que responde a la siguiente ecuación.

$$X_t = T_t \times E_t \times I_t$$

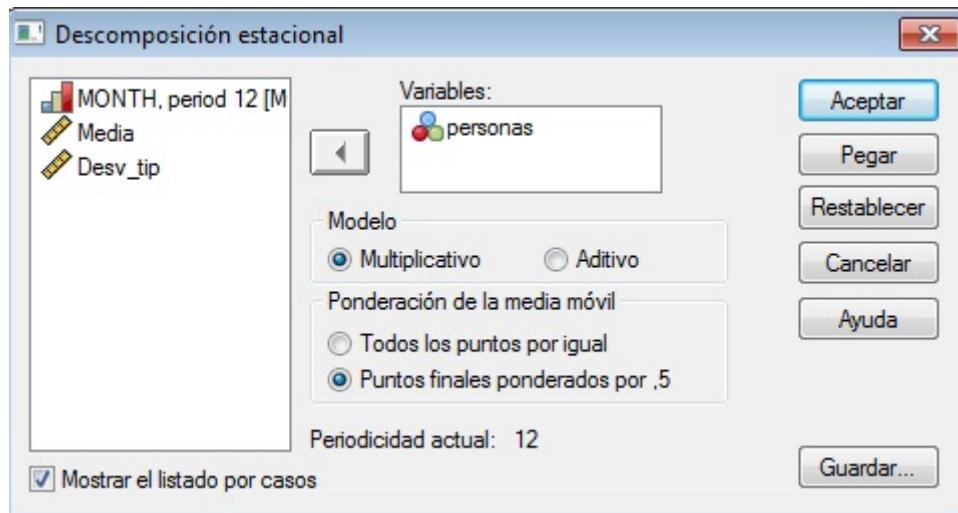
Donde T_t es la componente de tendencia, E_t es la componente estacional I_t es la componente irregular de la serie.

3. Extrae las componentes de la serie (Tendencia-ciclo, Estacionalidad, e Irregular) . Indica los 3 primeros valores de cada componente.

Para realizar la extracción de las componentes de la serie realizaremos la desestacionalización con la opción : Analizar ->Series Temporales ->Descomposición Estacional.

The screenshot shows the SPSS interface with the title bar "trabajo_hoteles.sav - SPSS Editor de datos". The menu bar is visible with "Analizar" selected. A data table titled "13 : Media" is shown on the left, containing columns for "personas" and "MONTH". The "Analizar" menu is expanded, showing sub-options like "Informes", "Estadísticos descriptivos", "Tablas", and "Series temporales". Under "Series temporales", "Descomposición estacional..." is highlighted. A context menu on the right lists options such as "Crear modelos...", "Aplicar modelos...", "Suavizado exponencial...", "Autorregresión...", "ARIMA...", "Descomposición estacional...", "Análisis espectral...", "Gráficos de secuencia...", "Autocorrelaciones...", and "Correlaciones cruzadas...".

Seleccionaremos como variable *personas* marcaremos en la casilla *multiplicativo* y puntos finales ponderados por 0.5 ya que es un periodo par (12 meses)



Y se nos generarán las siguientes 4 variables en nuestro Editor de Datos .

ERR_1	SAS_1	SAF_1	STC_1	YEAR	MONTH	DATE
1,01786	81427,29226	,68351	79998,15250	2004	1	JAN 2004
1,03470	82198,15466	,86832	79441,28258	2004	2	FEB 2004
.95367	74698,40082	1,03817	78327,54273	2004	3	MAR 2004
,99628	77554,72567	1,08130	77844,14748	2004	4	APR 2004
1,03694	79919,62922	1,08759	77072,64948	2004	5	MAY 2004
1,00366	76498,93560	1,10382	76219,71177	2004	6	JUN 2004
,94422	71089,23432	1,15850	75288,79611	2004	7	JUL 2004
1,01628	76908,14640	1,20472	75676,33701	2004	8	AUG 2004
1,01770	77597,66879	1,13885	76248,07313	2004	9	SEP 2004
,99466	76489,85208	1,01922	76900,23867	2004	10	OCT 2004
,97575	75554,42055	,89500	77432,01430	2004	11	NOV 2004
1,00990	79420,26671	,72100	78641,38799	2004	12	DEC 2004

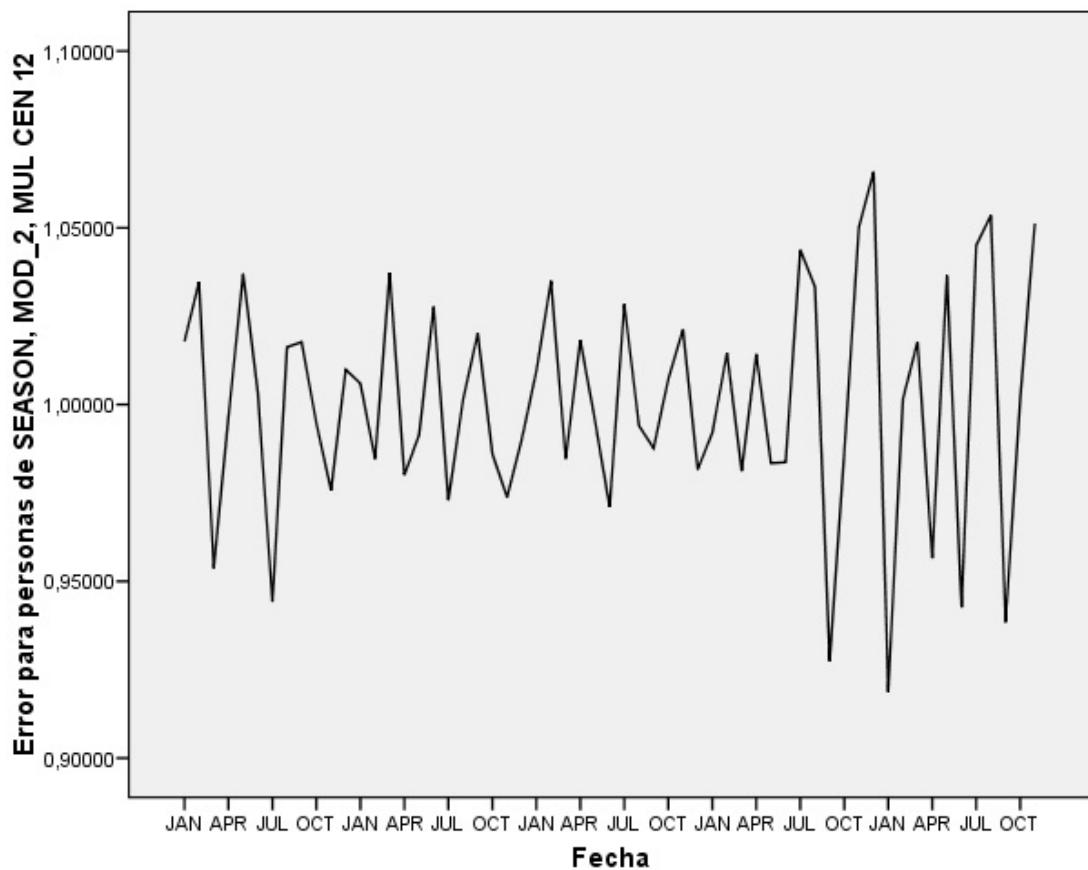
SAF_1 Es la componente estacional estimada.

STC_1 es la componente tendencia estimada.

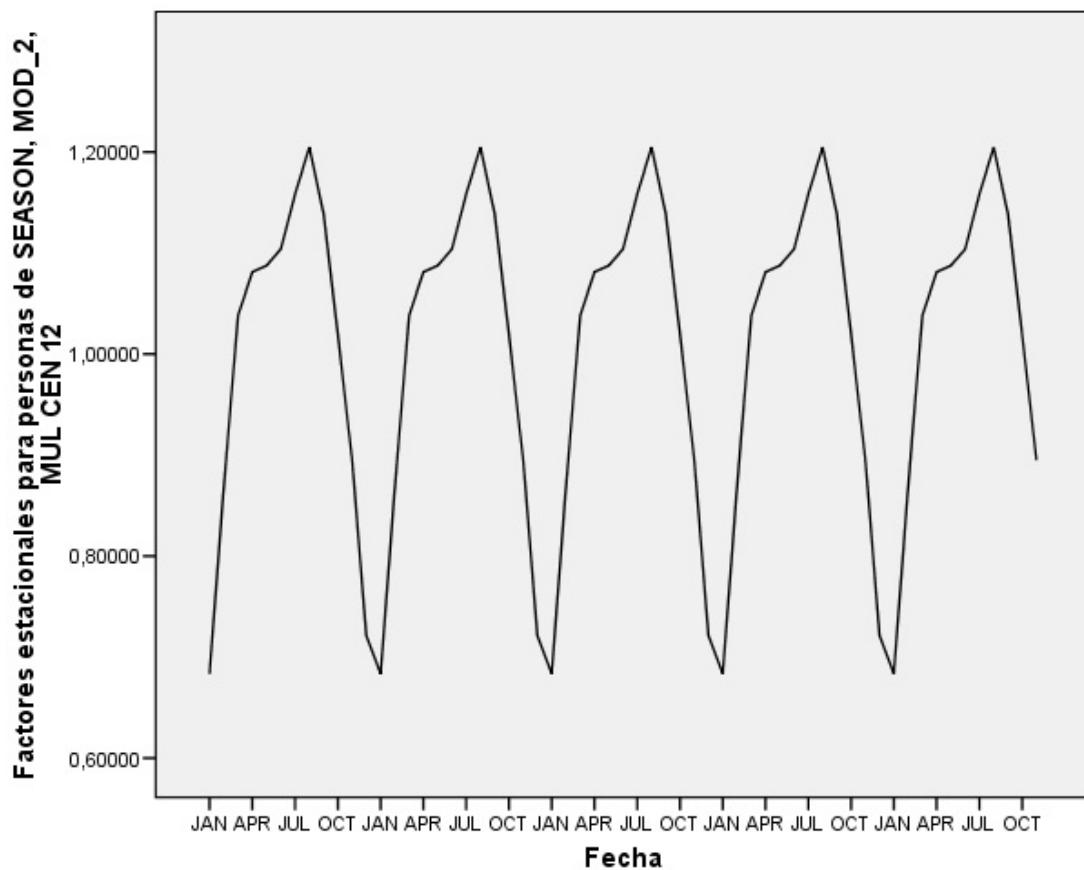
ERR_1 es la componente irregular estimada.

SAS_1 es la serie desestacionalizada.

Para aclarar más hemos realizado la representación individual de las componentes de la serie , usando la opción Anализar->Series Temporales-> Gráficos de secuencia.

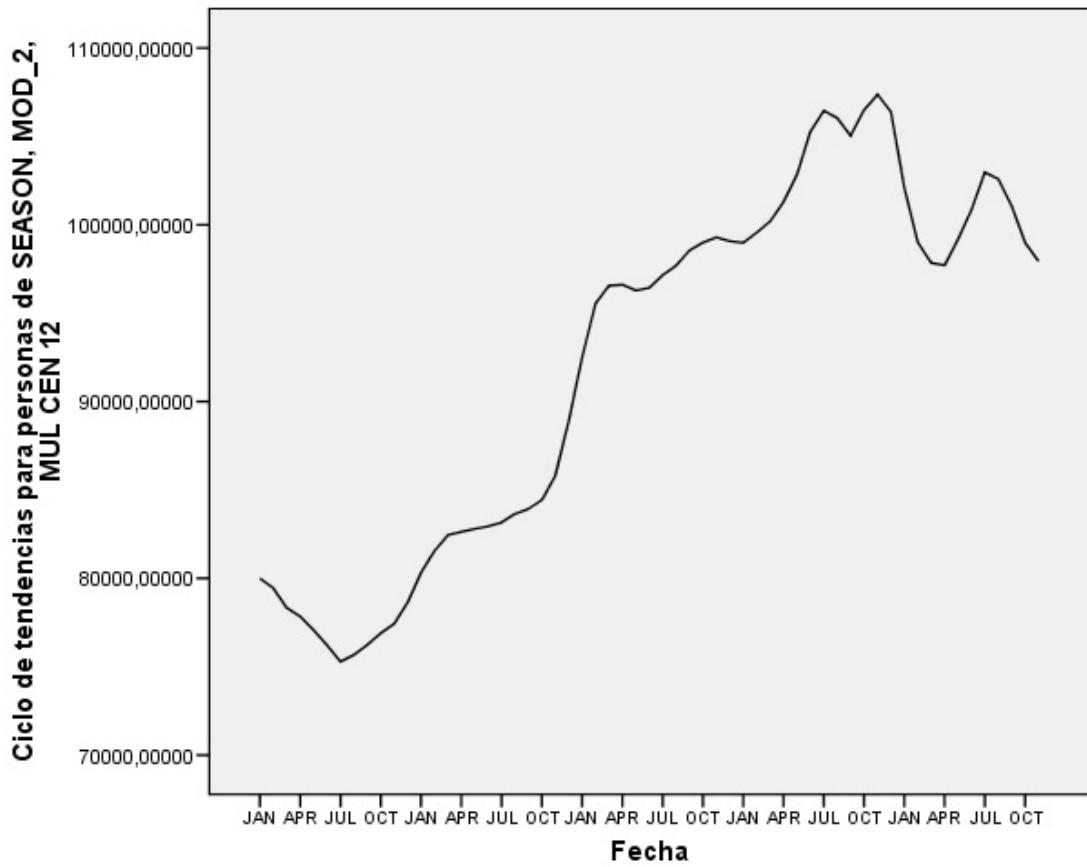


ERR_1 es la componente irregular estimada, al ser un modelo multiplicativo, dicha componente oscilará entorno al valor 1, si fuera un modelo aditivo, oscilaría entorno al valor 0.



SAF_1	DATE
,68351	JAN
,86832	FEB
1,03817	MAR
1,08130	APR
1,08759	MAY
1,10382	JUN
1,15850	JUL
1,20472	AUG
1,13885	SEP
1,01922	OCT
,89500	NOV
,72100	DEC

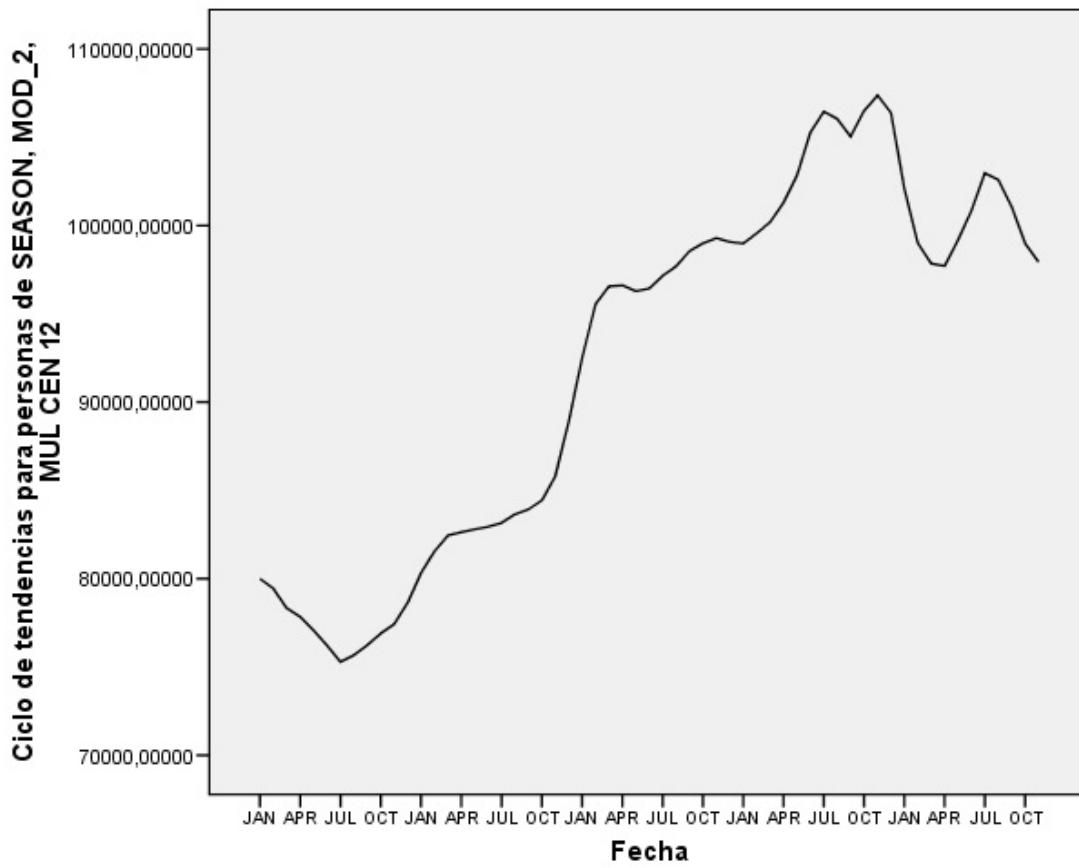
SAF_1 Es la componente estacional estimada, repetida cada 12 meses con los valores anteriormente indicados.



STC_1 es la componente tendencia estimada.

4. ¿Cómo dirías que es la tendencia? Intenta dar una explicación a la forma que presenta.

Para resolver esta cuestión pincharemos en Analizar series temporales. -> Gráficos de secuencia, y pasarámos como variable dependiente la tendencia. Devolviéndonos el siguiente grafico .



Concluimos que la tendencia es creciente hasta finales de 2007 y decreciente en adelante, parece tomar la forma de un modelo cuadrático o cubico, aunque para verificarlo lo veremos más adelante.

**5. Proporciona un modelo determinista que nos permita realizar predicciones, justificando qué curva has seleccionado para representar la tendencia.
Expresa el modelo de la forma:**

Ocupación Hotelera Estimada =

Realizamos la regresión lineal pinchando en Anализar -> Regresión lineal -> Estimación curvilínea, para decidir cuál es el modelo matemático que más se ajusta a nuestro modelo de estudio.

trabajo_hoteles.sav - SPSS Editor de datos

Archivo Edición Ver Datos Transformar Analizar Gráficos Utilidades Ventana ?

1 : YEAR_

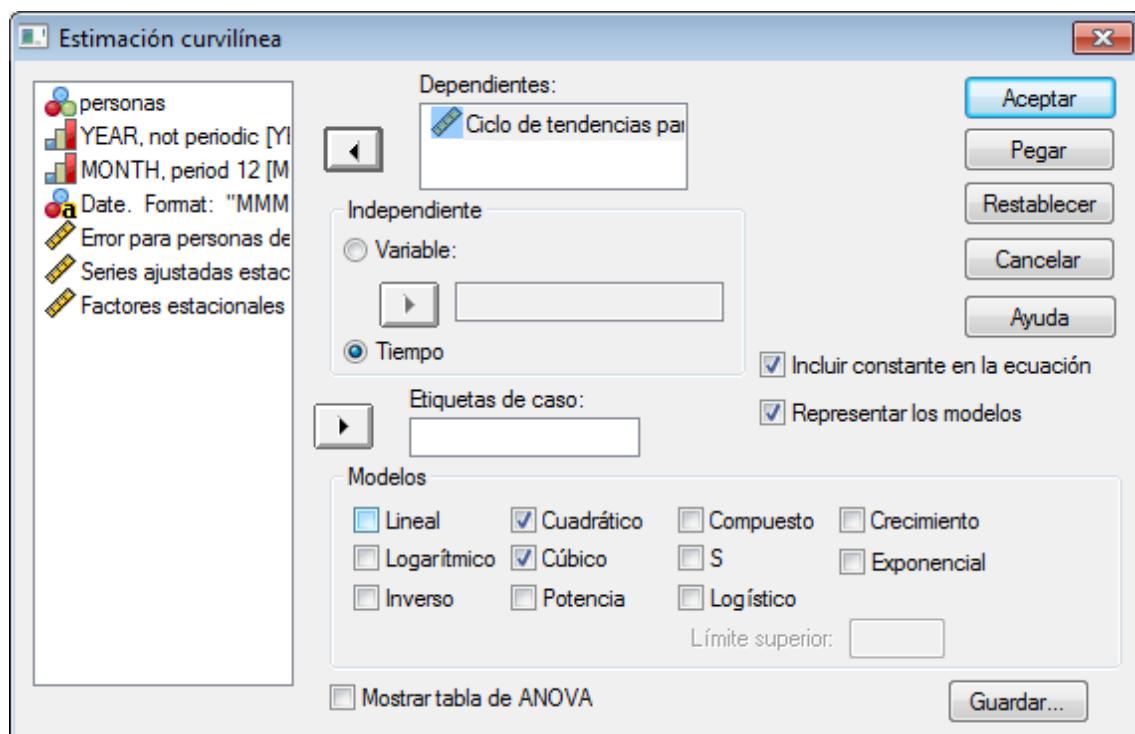
	personas	Media
1	55656	61189,40
2	71374	80157,80
3	77550	94098,80
4	83860	97992,60
5	86920	100444,0
6	84441	100226,0
7	82357	109047,6
8	92653	114551,8
9	88372	103135,0
10	77960	94509,80
11	67621	85245,20
12	57262	68147,50
13	55232	-
14	69725	-
15	88786	-
16	87578	-
17	89278	-
18	94092	-
19	93741	-
20	100909	-
21	97512	-

Analizar ▾

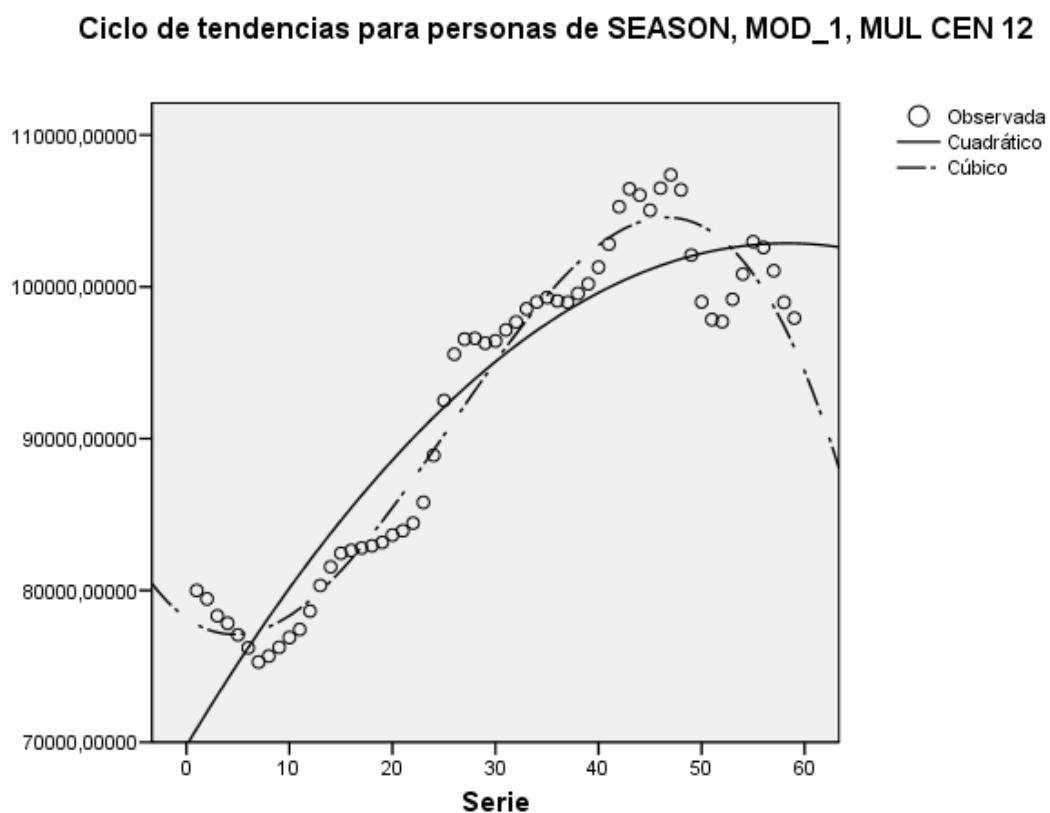
- Informes
- Estadísticos descriptivos
- Tablas
- Comparar medias
- Modelo lineal general
- Modelos lineales generalizados
- Modelos mixtos
- Correlaciones
- Regresión ▾
- Loglineal
- Clasificar
- Reducción de datos
- Escalas
- Pruebas no paramétricas
- Series temporales
- Supervivencia
- Respuesta múltiple
- Análisis de valores perdidos...
- Muestras complejas
- Control de calidad
- Curva COR...
- Lineal...
- Estimación curvilinea...
- Logística binaria...
- Logística multinomial ...
- Ordinal...
- Probit...
- No lineal...
- Estimación ponderada...
- Mínimos cuadrados en dos fases...
- Escalamiento óptimo...

	SAF 1	STC 1
9226	,68351	79998,15250
5466	,86832	79441,28258
0082	1,03817	78327,54273
2567	1,08130	77844,14748
2922	1,08759	77072,64948
1934	1,13030	83102,77311
7498	1,20472	83651,17839
1824	1,13885	83930,42392

Marcamos la tendencia como variable dependiente y seleccionamos los modelos que creemos que pueden corresponderse con nuestra variable de estudio. TAMBIÉN DEBERÍAMOS SELECCIONAR “MOSTRAR TABLA ANOVA” PARA DETERMINAR SI TODOS LOS TÉRMINOS DE LA CURVA MODELIZADA SON SIGNIFICATIVOS.



Devolviéndonos SPSS el siguiente gráfico:



Observamos que el modelo cúbico se ajusta mucho más a nuestro modelo que cualquier otro, aunque intuitivamente observando y razonando el contexto evolutivo y social del modelo de estudio(ocupación hotelera) creemos que es posible que la evolución se asemeje más a un modelo cuadrático que a un cúbico, no obstante nosotros realizaremos el estudio con ambos modelos.

SPSS nos devuelve los siguientes valores de los coeficientes para el modelo cuadrático y cúbico:

Resumen del modelo y estimaciones de los parámetros									
Ecuación	Resumen del modelo					Estimaciones de los parámetros			
	R cuadrado	F	gl1	gl2	Sig.	Constante	b1	b2	b3
Cuadrático	,877	200,508	2	58	,000	69763,893	1138,023	-9,747	
Cúbico	,954	382,238	3	55	,000	78186,100	-480,868	57,061	-,742

Con los cuales obtenemos nuestras ecuaciones ajustadas para cada modelo.

Cuadrático:

$$T_t = 69763.893 + 1138.023 t - 9.747 t^2$$

Cúbico:

$$T_t = 78186.100 - 480.868 t + 57.061 t^2 - 0.742 t^3$$

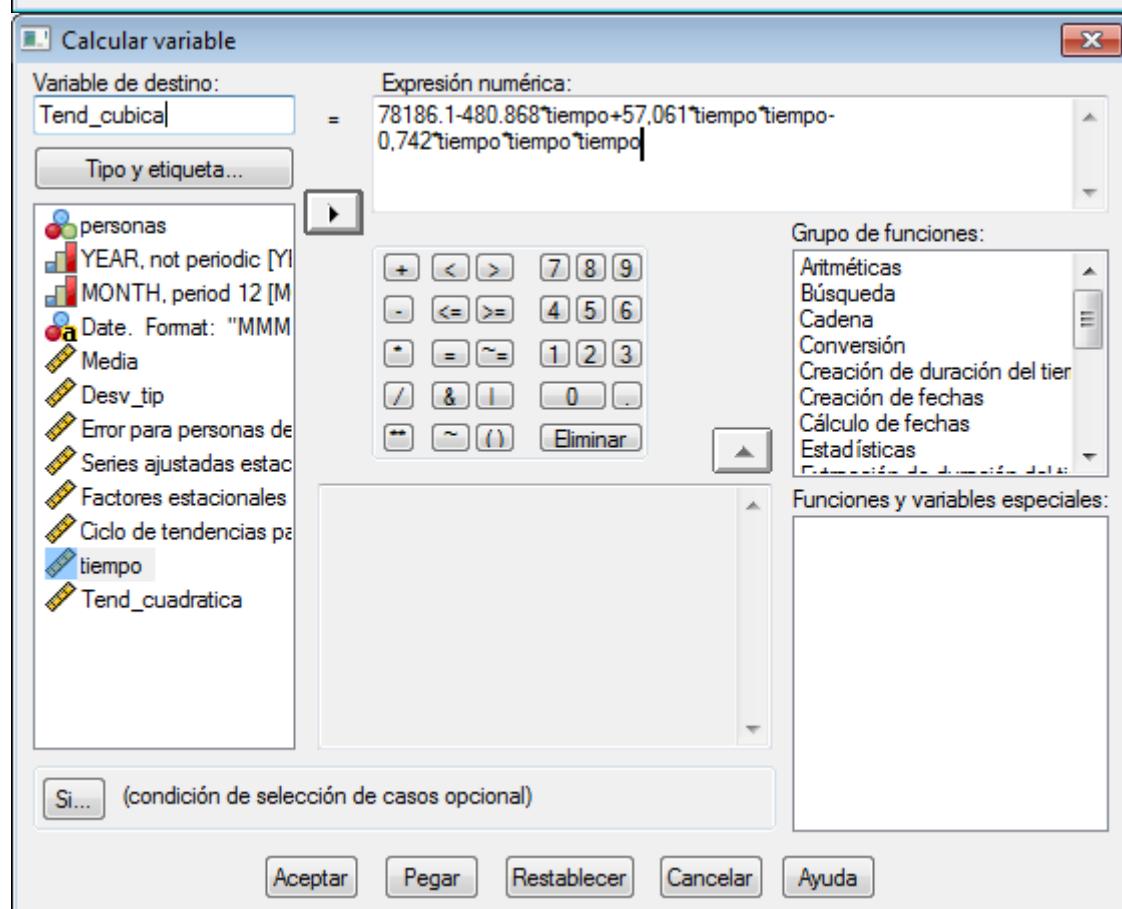
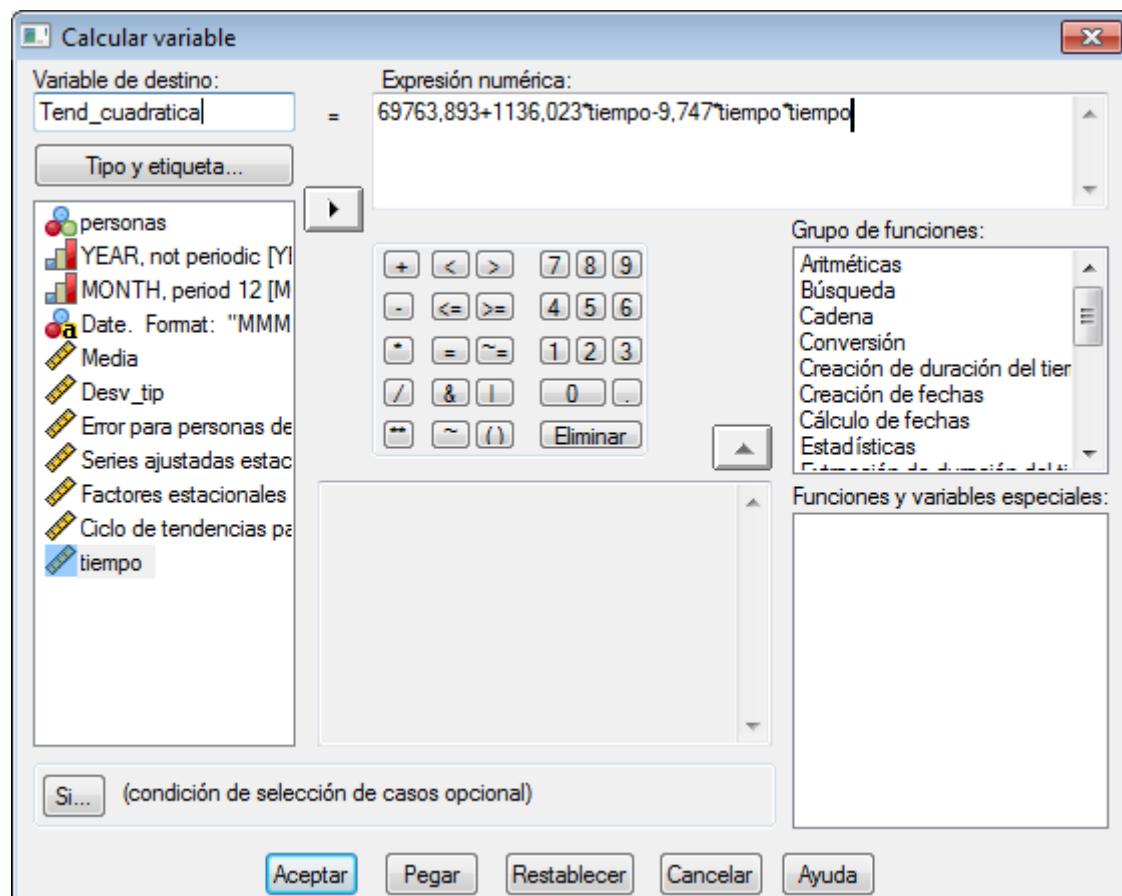
- 6. Determina los residuos del modelo anterior, es decir, la diferencia entre el valor observado y el estimado por el modelo determinista. Indica los 3 primeros residuos.**

Para resolver a esta cuestión tendremos que calcular la ecuación con el modelo ajustado, pincharemos en Transformar -> Calcular Variable

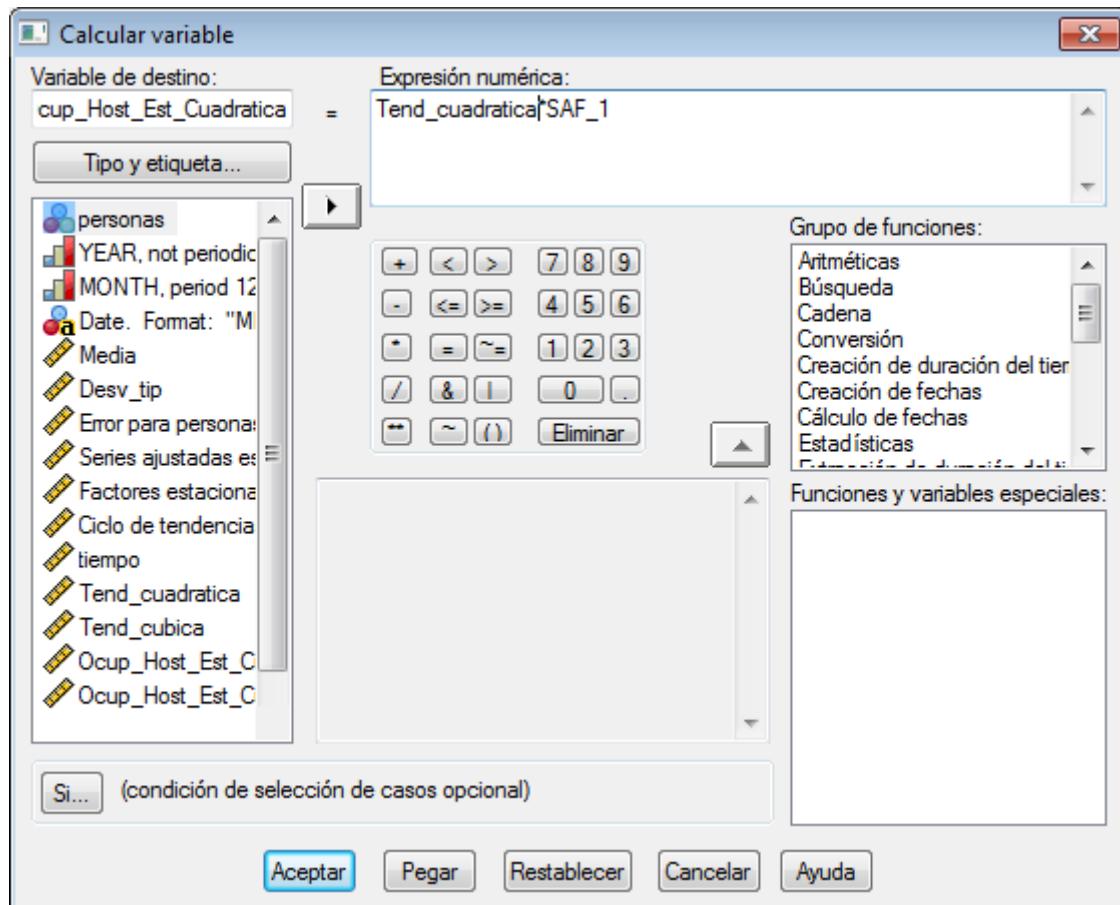
Primero creamos una variable denominada “tiempo” usando el comando:

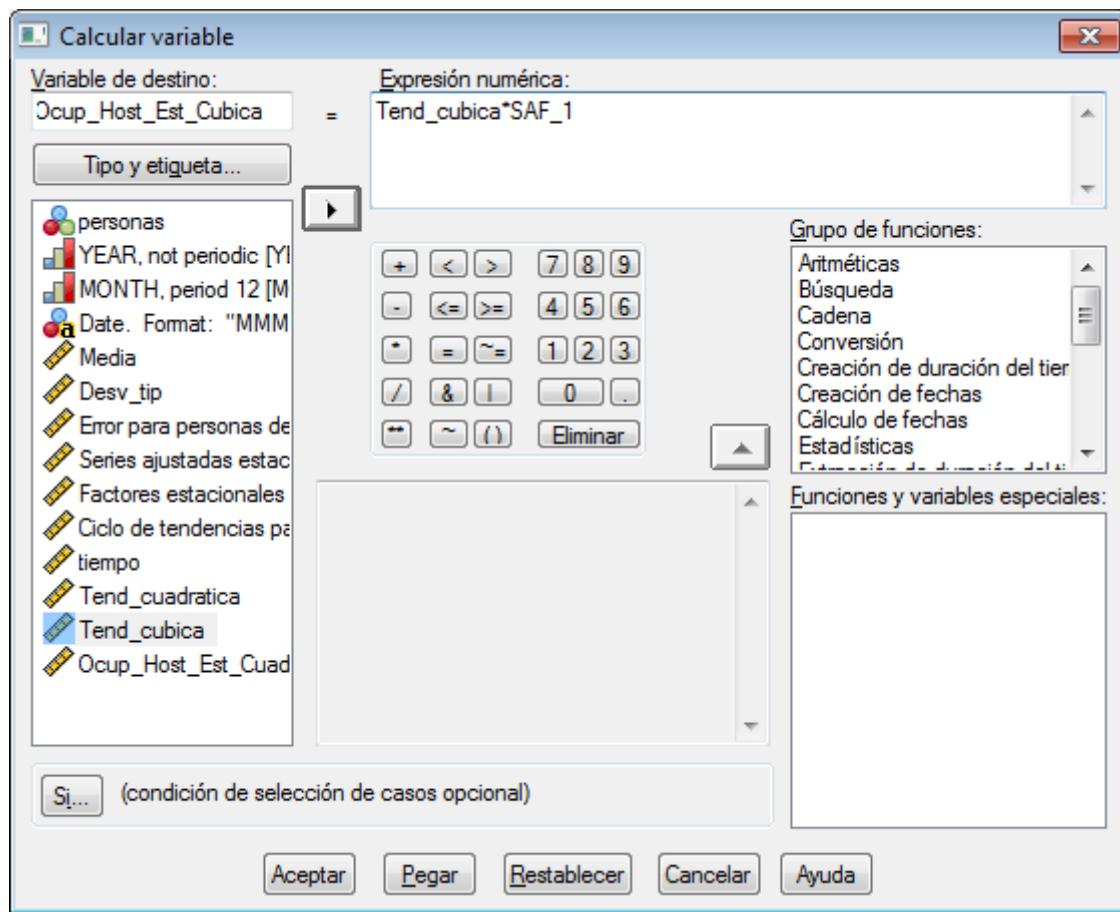
Tiempo = \$casenum

A continuación, introducimos la transformación correspondiente a la tendencia cuadrática y cúbica.



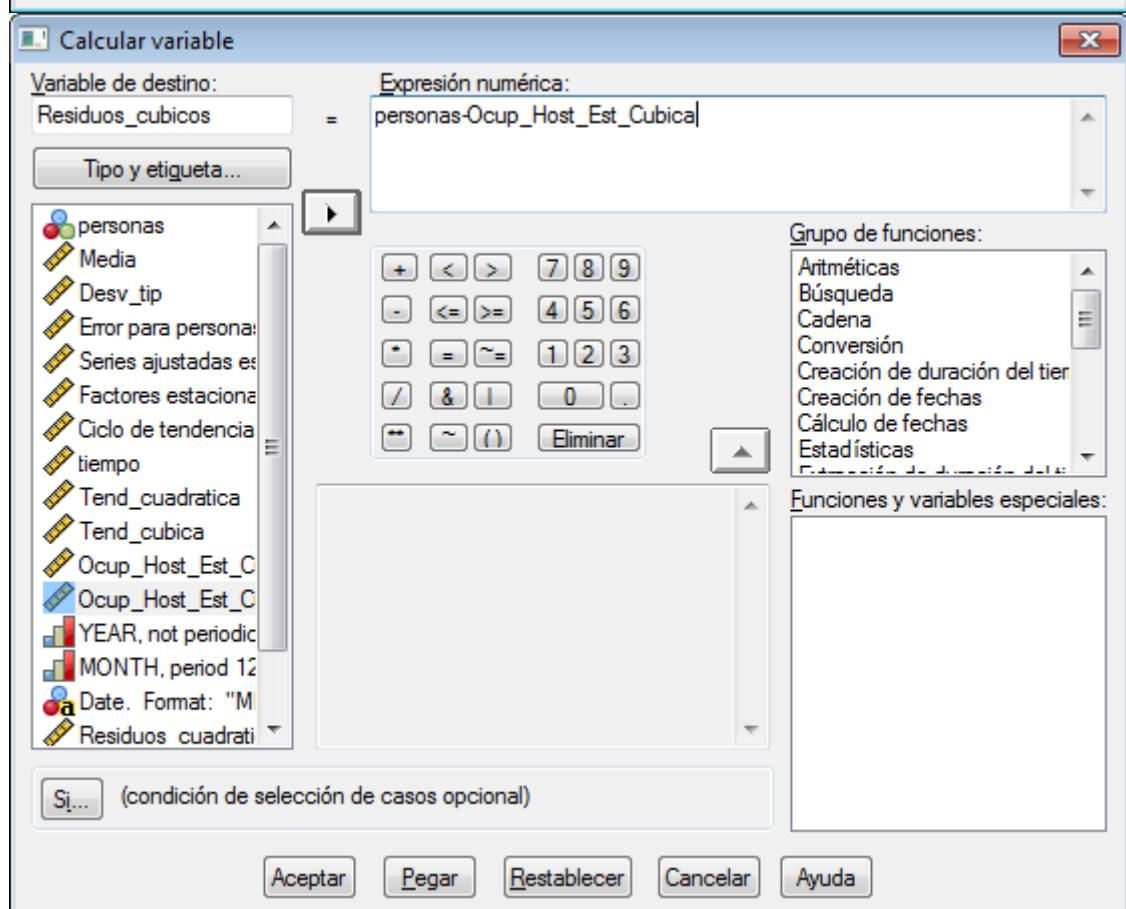
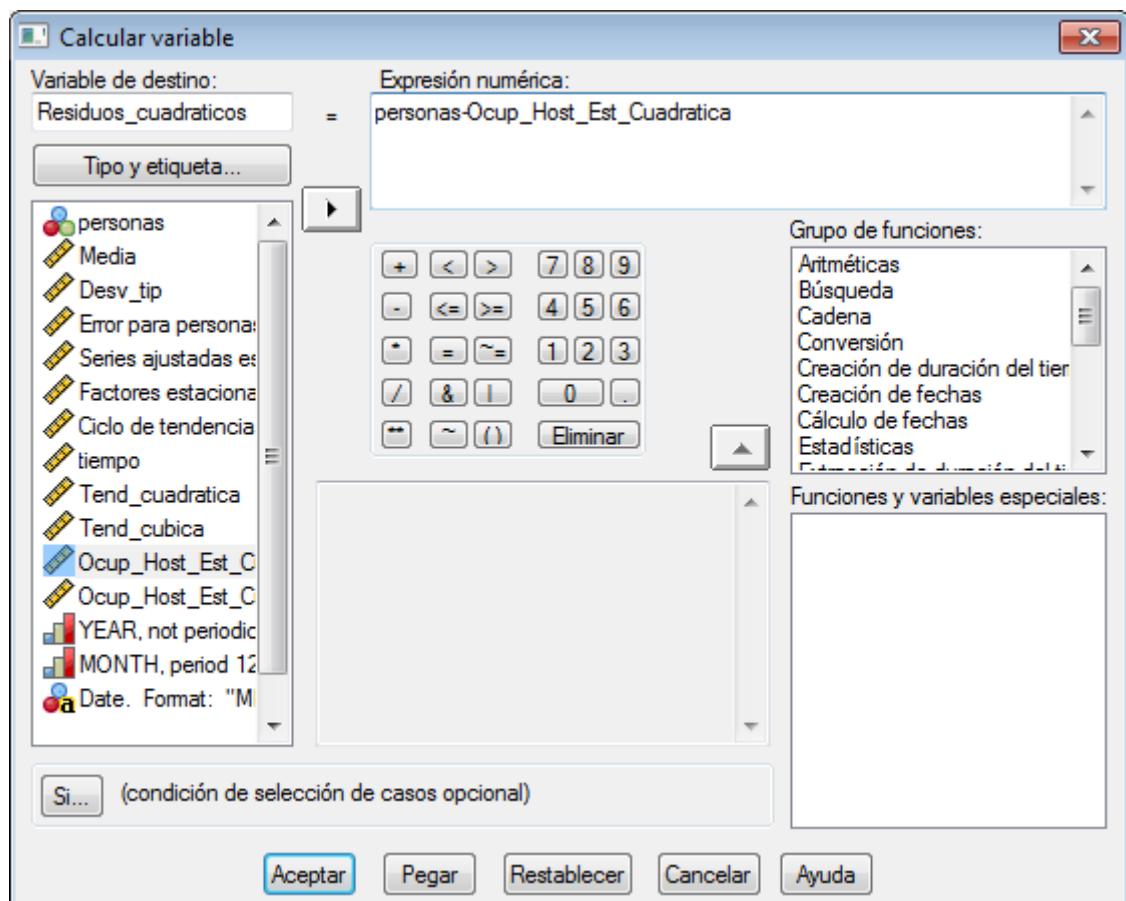
Posteriormente definiremos la ecuación obtenida en la estimación curvilínea y crearemos la variable "Ocupación hotelera Estimada" que es la ecuación de nuestro modelo ajustado (cuadrático y cúbico).





Seguidamente tendremos que volver a definir otra variable a la que llamaremos residuos y será la resta de nuestra variable a estudio "personas" y "Ocupación hotelera Estimada" tal que:

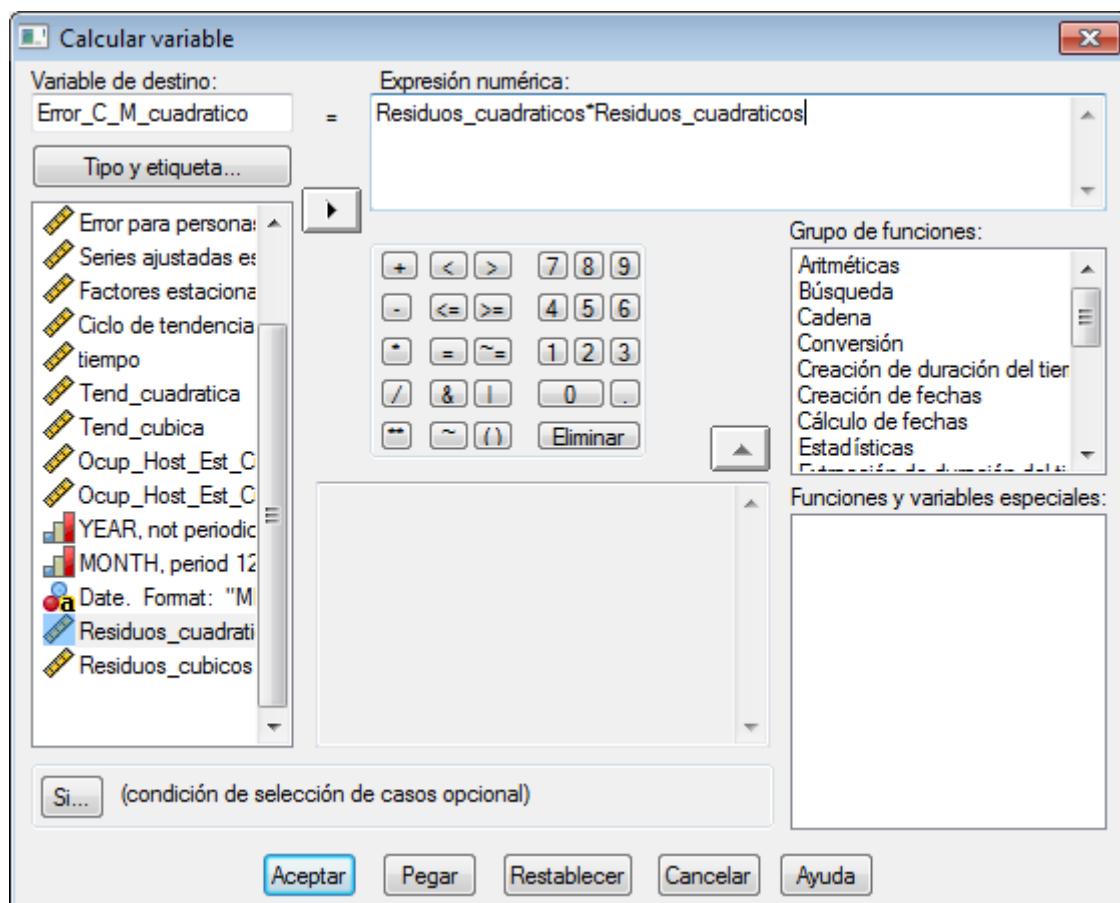
Residuos = Personas – Ocupación hotelera Estimada.

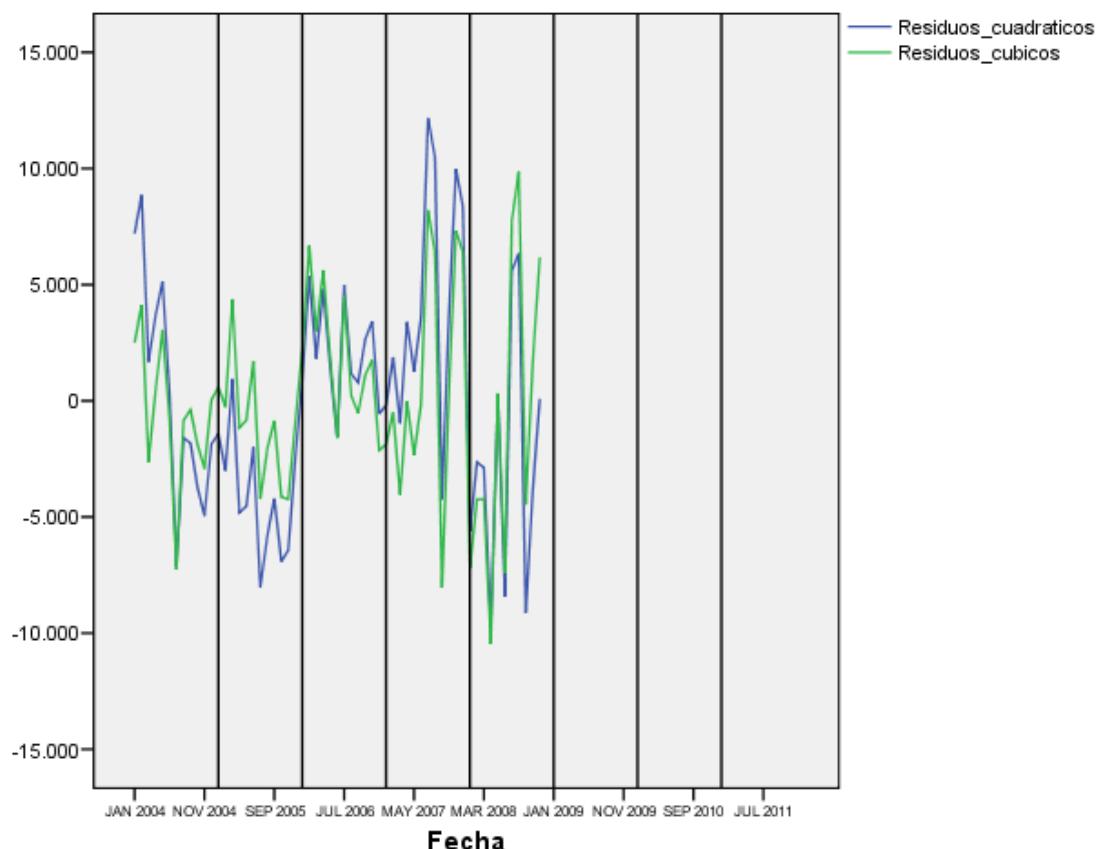
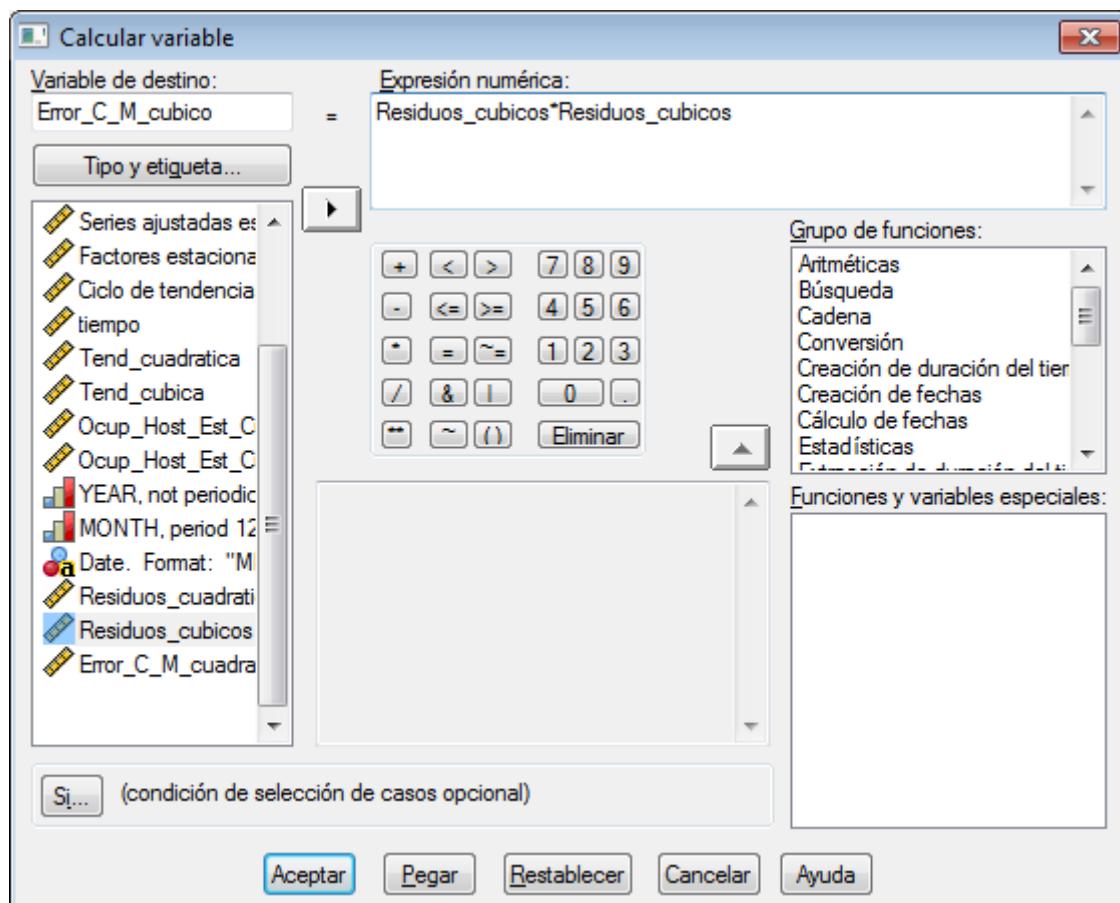


Los tres primeros valores que nos devuelve para cada modelo propuesto son los siguientes:

	YEAR	MONTH	DATE	Residuos_cuadraticos	Residuos_cubicos	Error_C_M_cuadratico	Error_C_M_cubico
1	2004	1	JAN 2004	7202,18	2505,55	51871417,34	6277801,94
2	2004	2	FEB 2004	8857,87	4125,79	78461891,48	17022143,00
3	2004	3	MAR 2004	1675,78	-2635,52	2808245,93	6945972,02
4	2004	4	APR 2004	3679,33	461,29	13537505,18	212791,65
5	2004	5	MAY 2004	5132,67	3049,71	26344319,83	9300744,60
6	2004	6	JUN 2004	297,81	-768,13	88688,82	590017,87
7	2004	7	JUL 2004	-7123,88	-7266,44	50749638,90	52801130,91
8	2004	8	AUG 2004	-1590,38	-846,93	2529314,39	717293,33
9	2004	9	SEP 2004	-1823,21	-389,08	3324096,03	151382,28
10	2004	10	OCT 2004	-3729,91	-1887,26	13912217,39	3561740,09
11	2004	11	NOV 2004	-4946,06	-2916,72	24463487,96	8507231,96
12	2004	12	DEC 2004	-1854,65	50,45	3439736,35	2545,38
13	2005	1	JAN 2005	-1420,32	587,14	2017298,57	344730,85
14	2005	2	FEB 2005	-3003,27	-262,91	9019619,43	69120,52
15	2005	3	MAR 2005	944,82	4374,54	892684,37	19136621,61
16	2005	4	APR 2005	-4813,79	-1154,20	23172599,50	1332182,14
17	2005	5	MAY 2005	-4537,09	-836,15	20585177,38	699144,13
18	2005	6	JUN 2005	-2000,19	1712,38	4000766,26	2932229,66
19	2005	7	JUL 2005	-8009,82	-4221,04	64157142,17	17817157,14
20	2005	8	AUG 2005	-5812,04	-2043,17	33779789,89	4174534,61
21	2005	9	SEP 2005	-4212,18	-861,87	17742466,68	742824,05
22	2005	10	OCT 2005	-6912,41	-4145,09	47781453,18	17181759,06

Como complemento a esta cuestión nos hemos propuesto la comparación de los errores cuadráticos medios usando ambas los modelos cuadrático y cúbico. FALTARÍA HACER LA MEDIA DE LAS VARIABLES “ERROR_C_M”, para comparar los valores numéricos.





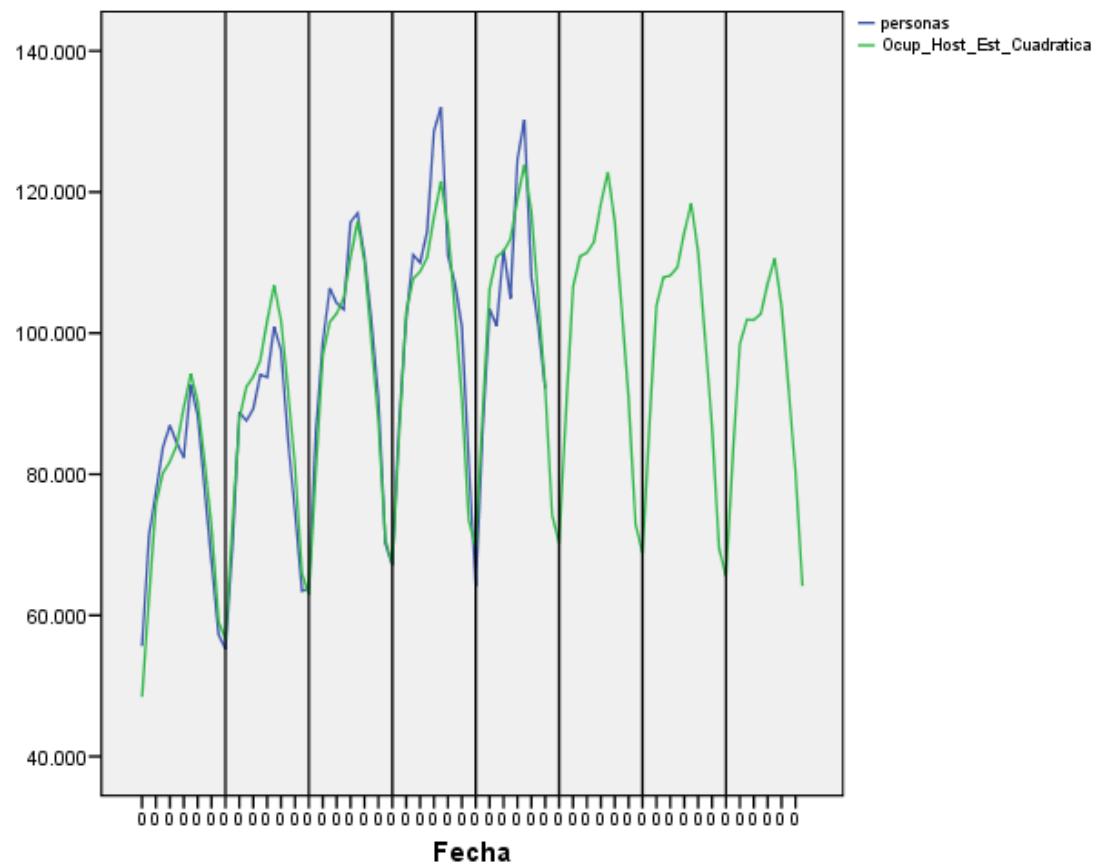
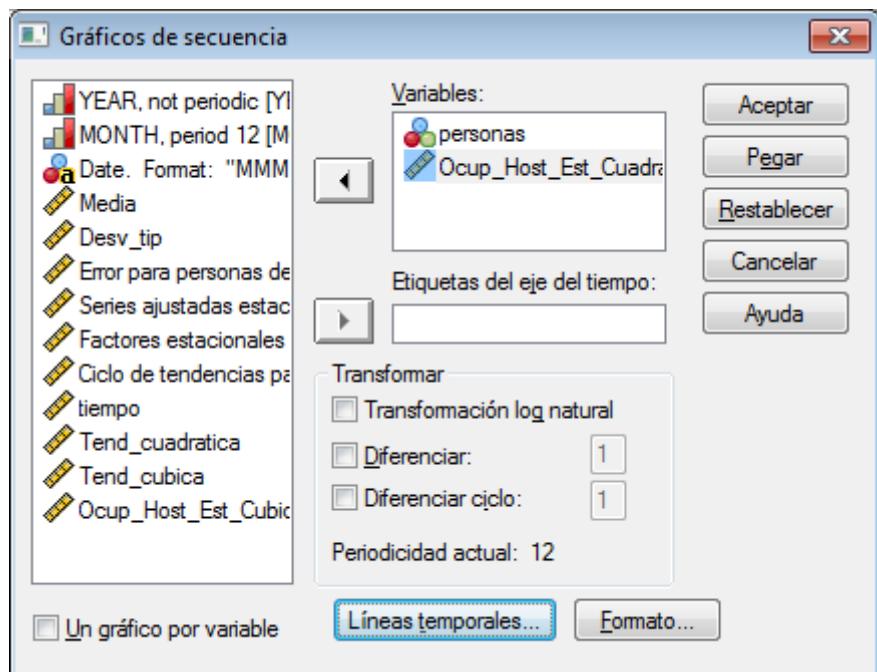
7. Determina las predicciones de la ocupación hotelera para el próximo año 2009. (Indica sólo las predicciones para los meses de enero, agosto y diciembre).

Para realizar las predicciones tendremos que completar la columna de la variable estacional SAF_1 hasta llegar al mes de predicción deseado y volver a calcular nuestra variable “Ocupación hotelera” incluyendo los últimos meses de predicción que no estaban incluidos inicialmente.

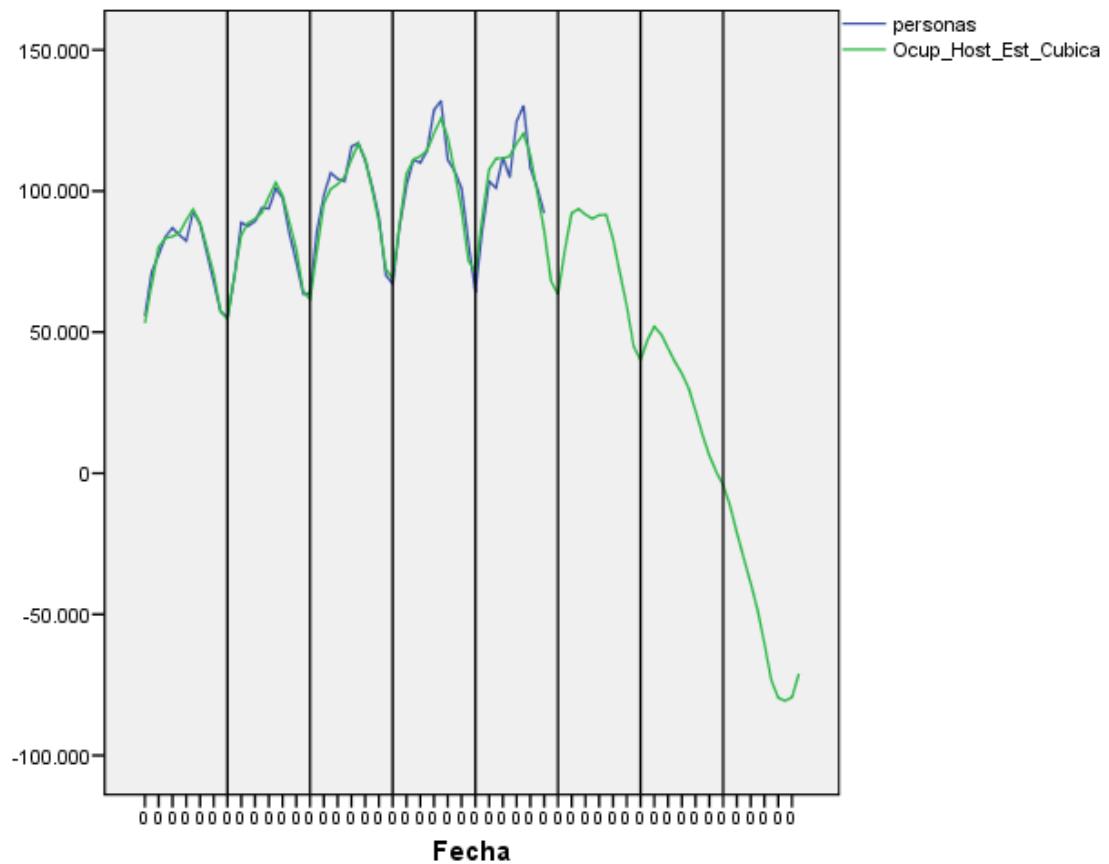
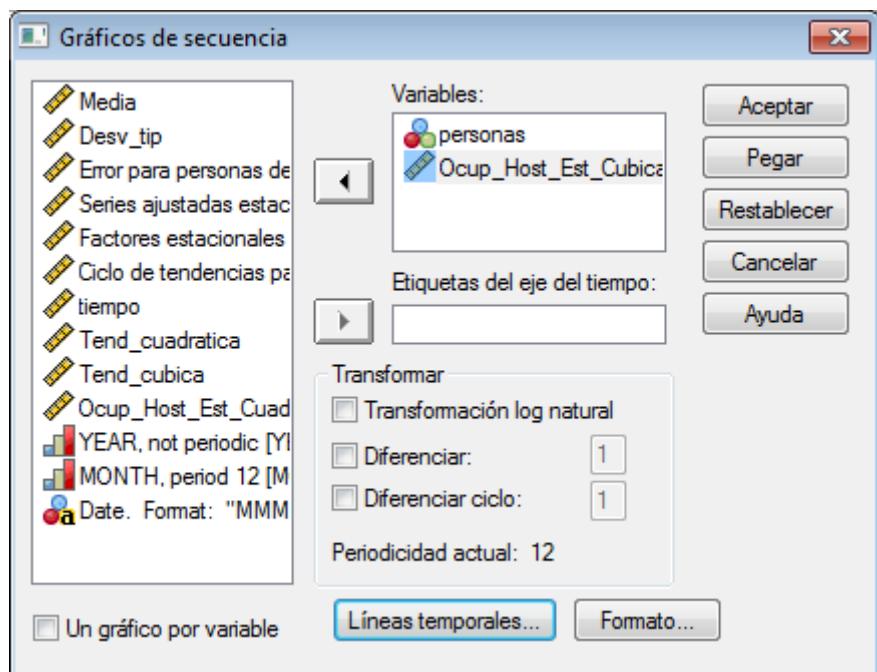
Ocup_Host_Est_Cuadratica	Ocup_Host_Est_Cubica	YEAR	MONTH	DATE
104841,39	99349,36	2008	10	OCT 2008
92059,37	85967,18	2008	11	NOV 2008
74144,79	68121,23	2008	12	DEC 2008
70259,38	63400,08	2009	1	JAN 2009
89202,01	78908,59	2009	2	FEB 2009
106566,05	92222,78	2009	3	MAR 2009
110882,71	93663,97	2009	4	APR 2009
111395,93	91619,52	2009	5	MAY 2009
112902,48	90164,92	2009	6	JUN 2009
118309,85	91462,12	2009	7	JUL 2009
122813,46	91593,39	2009	8	AUG 2009
115871,03	83044,79	2009	9	SEP 2009
103476,53	70956,31	2009	10	OCT 2009
90651,50	59176,33	2009	11	NOV 2009
72841,97	45003,23	2009	12	DEC 2009
68864,41	39992,04	2010	1	JAN 2010
87226,74	47229,20	2010	2	FEB 2010
103961,52	51968,83	2010	3	MAR 2010
107917,05	49203,82	2010	4	APR 2010
108158,59	44293,78	2010	5	MAY 2010
109358,62	39428,35	2010	6	JUN 2010

A continuación veremos la efectividad de las predicciones a largo plazo(3 años) de ambos modelos para saber cuál de los 2 realiza unas mejores estimaciones.

Modelo Cuadrático:



Modelo cúbico:



Como podemos comprobar el modelo cubico a partir del primer año tiene una pendiente muy decreciente que pronostica un descenso súbito de la demanda hostelería, con lo cual desecharíamos esta opción y nos decantaríamos por el modelo cuadrático que se asemejaría mas a la realidad en el sector hostelería.

Nos proponemos ahora analizar la serie del fichero hoteles_ST.sav mediante una técnica de alisado exponencial.

8. ¿Qué técnica de alisado exponencial te parece más adecuada para analizar esta serie? Razona tu respuesta.

Usaremos la técnica de Holt-Winters multiplicativo, ya que la serie presenta componente estacional y dicha componente parece ir MULTIPLICADA por la componente tendencia (ver apartado 2).

9. Aplica la técnica de alisado que has considerado adecuada, seleccionando como parámetros de alisado aquellos que minimizan la suma de cuadrados de los errores (búsqueda en rejilla con precisión de 0.1). Determina cómo quedarían las fórmulas recurrentes de las series alisadas e interpreta el significado de los parámetros de alisado obtenidos.

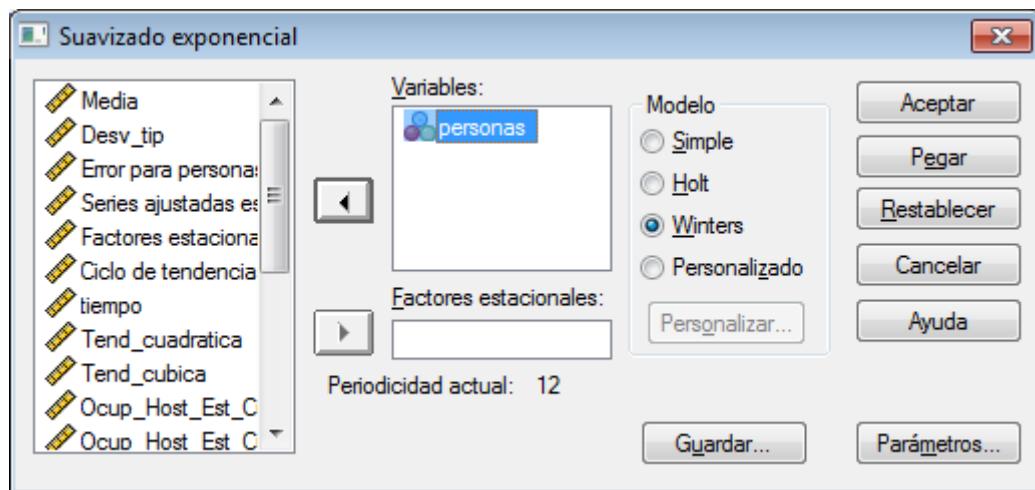
Tendremos que seleccionar el menú Analizar -> Series Temporales-> suavizado exponencial. Seleccionaremos Winters y marcaremos en cada parámetro la rejilla de búsqueda definida en el enunciado de 0.1

The screenshot shows the SPSS interface with the following details:

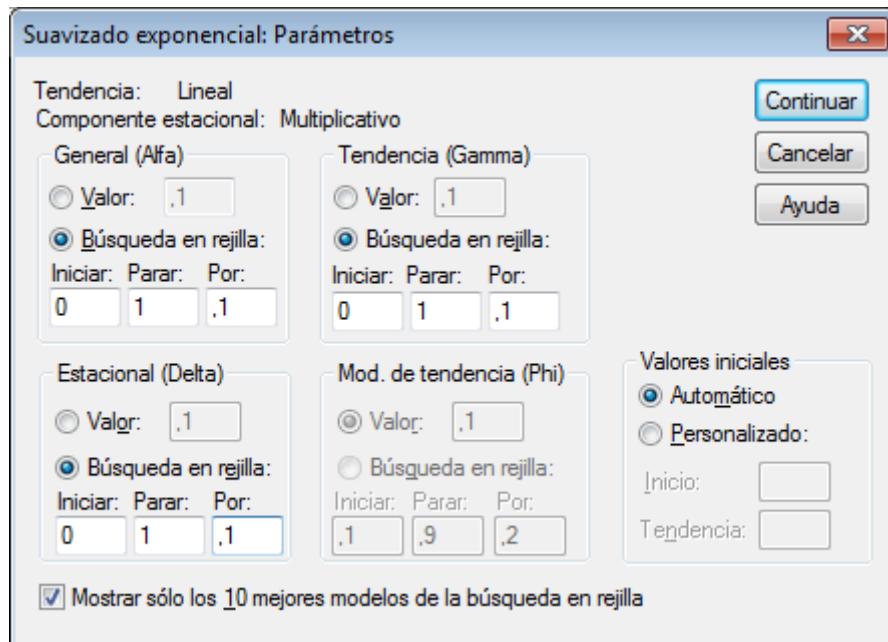
- Menu Bar:** Archivo, Edición, Ver, Datos, Transformar, **Analizar**, Gráficos, Utilidades, Ventana, ?
- Submenu Analizar:** Informes, Estadísticos descriptivos, Tablas, Residuos, Error_cuadrático_medio, YEAR, MONTH, DATE, var, var, var, var.
- Submenu Series temporales:** Comparar medias, Modelo lineal general, Modelos finales generalizados, Modelos mixtos, Correlaciones, Regresión, Loglineal, Clasificar, Reducción de datos, Escalas, Pruebas no paramétricas.
- Submenu Suavizado exponencial...:** Crear modelos..., Aplicar modelos..., Suavizado exponencial..., Análisis de valores perdidos..., Autorregresión..., ARIMA..., Descomposición estacional..., Análisis espectral..., Gráficos de secuencia..., Autocorrelaciones..., Correlaciones cruzadas...
- Data View:** A table titled 'tiempo_1' with columns Est, Pred, tiempo_1, Residuos, Error_cuadrático_medio, YEAR, MONTH, DATE, var, var, var, var. The data spans rows 46 to 75, showing values for various dates from 2007 to 2009.
- Status Bar:** SPSS El procesador está preparado, Vista de datos, Vista de variables, 15:47, 16/12/2010.

Pasaremos la variable personas y seleccionaremos

Winters:



Pincharemos en parámetros y definiremos la rejilla de búsqueda:



SPSS nos muestra los modelos propuestos para el ajuste, de donde resaltamos 2 propuestas:

Sumas menores de los errores cuadráticos

Serie	Rango del modelo	Alpha (Nivel)	Gamma (Tendencia)	Delta (Estación)	Sumas de los errores cuadráticos
personas	1	,40000	,00000	,00000	1350061833
	2	,50000	,00000	,00000	1368574435
	3	,30000	,00000	,00000	1376042409
	4	,40000	,00000	,10000	1392337861
	5	,30000	,10000	,00000	1398119008
	6	,50000	,00000	,10000	1405710254
	7	,40000	,10000	,00000	1409213695
	8	,60000	,00000	,00000	1413669857
	9	,30000	,00000	,10000	1422941475
	10	,40000	,00000	,20000	1437131201

Parámetros del suavizado

Serie	Alpha (Nivel)	Gamma (Tendencia)	Delta (Estación)	Sumas de los errores cuadráticos	gl error
personas	,40000	,00000	,00000	1350061833	46

A continuación, se muestran los parámetros con las sumas menores de errores cuadráticos. Estos parámetros se utilizan para pronosticar.

Los valores iniciales de nivel tendencia así como de los índices estacionales quedan indicados a continuación:

Estado de suavizado inicial

	personas
Índices estacionales 1	68,63331
2	87,15135
3	103,91117
4	108,27082
5	109,16571
6	108,88687
7	115,99511
8	120,40955
9	114,06484
10	101,65562
11	89,52863
12	72,32902
Nivel Tendencia	72618,31
	758,42130

De todos los modelos propuestos analizaremos cual es el más apropiado para ajustar nuestro modelo, en nuestro caso concreto elegiremos el modelo propuesto 1 (por tener menor error cuadrático medio), cuyos parámetros de alisado son $\alpha=0.4$ $\gamma=0.0$ y $\delta=0.0$.

Estos parámetros pueden oscilar entre 0 y 1, significando 0, que a todas las observaciones les da el mismo peso, y 1 que las observaciones más influyentes para determinar el modelo son las más recientes.

El parámetro $\alpha = 0.4$, nos indica que para estimar el nivel de la serie en cada instante, se le da mayor peso a las observaciones finales que a las iniciales.

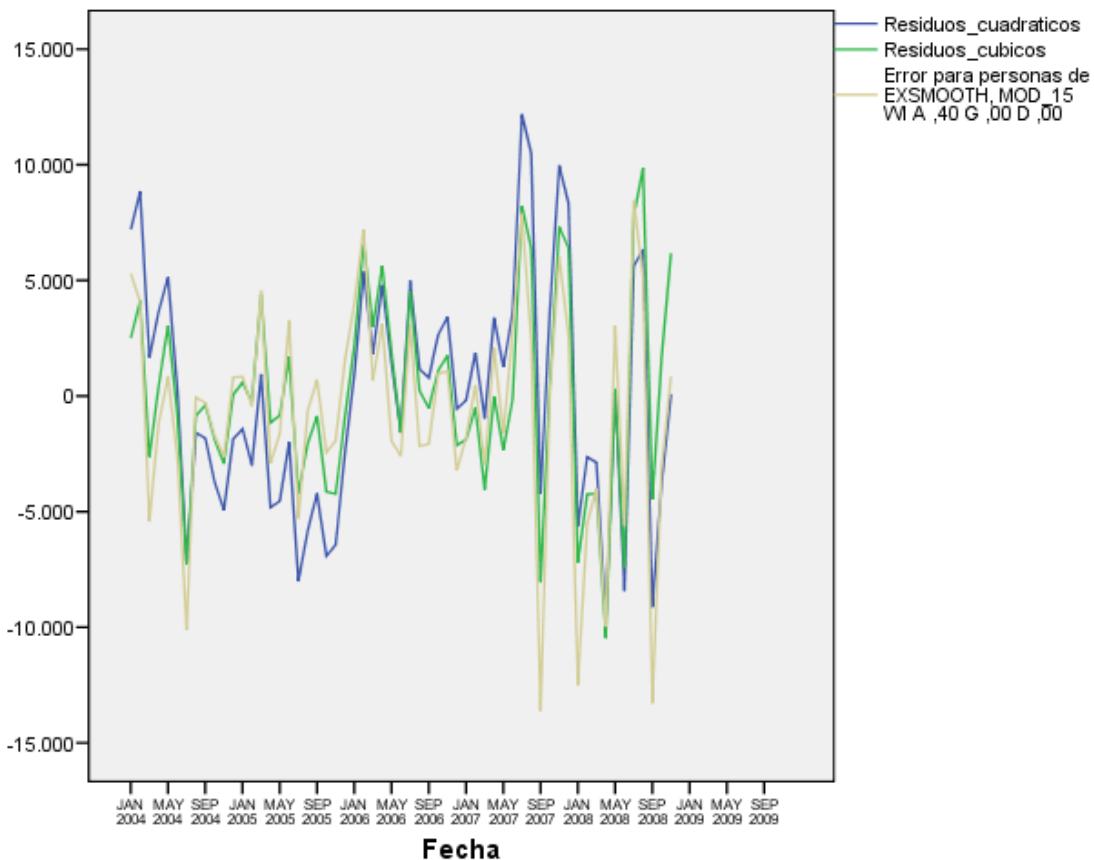
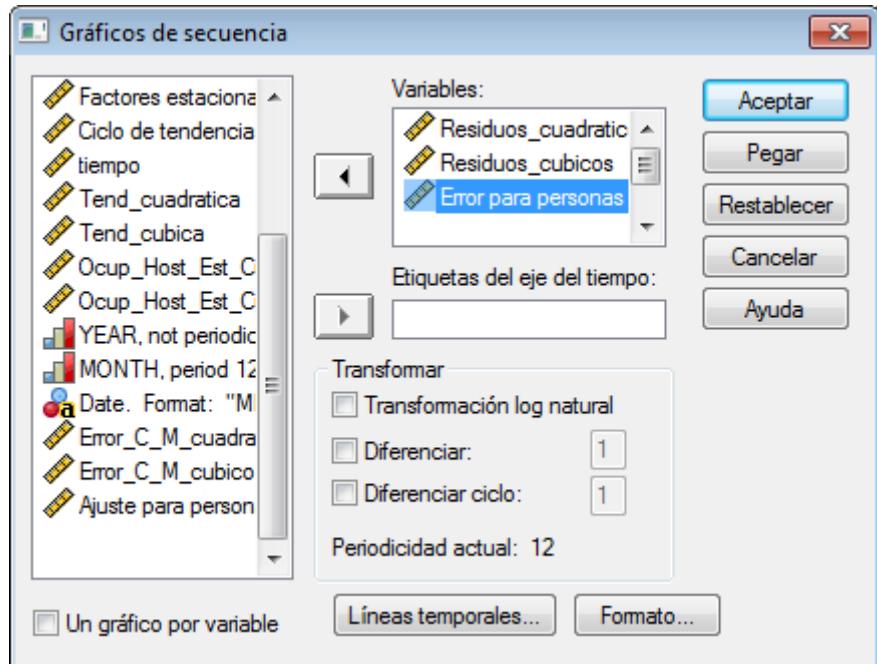
El parámetro $\gamma = 0$ nos indica que para estimar la pendiente de la serie en cada instante se le da el mismo peso a las observaciones iniciales que las observaciones finales en el ajuste de nuestro modelo.

El parámetro $\delta = 0$ se interpreta como que la componente estacional en el modelo no se modifica a lo largo del tiempo, y que las observaciones finales tienen el mismo peso que las iniciales para el cálculo de nuestro modelo alisado.

FALTARÍA ESCRIBIR LAS FÓRMULAS RECURRENTES RESULTANTES!

10. Compara los residuos obtenidos mediante esta técnica de alisado con los residuos de la descomposición clásica. Para ello, realiza un gráfico comparativo de los residuos: represéntalo en tu folio a mano alzada y coméntalo.

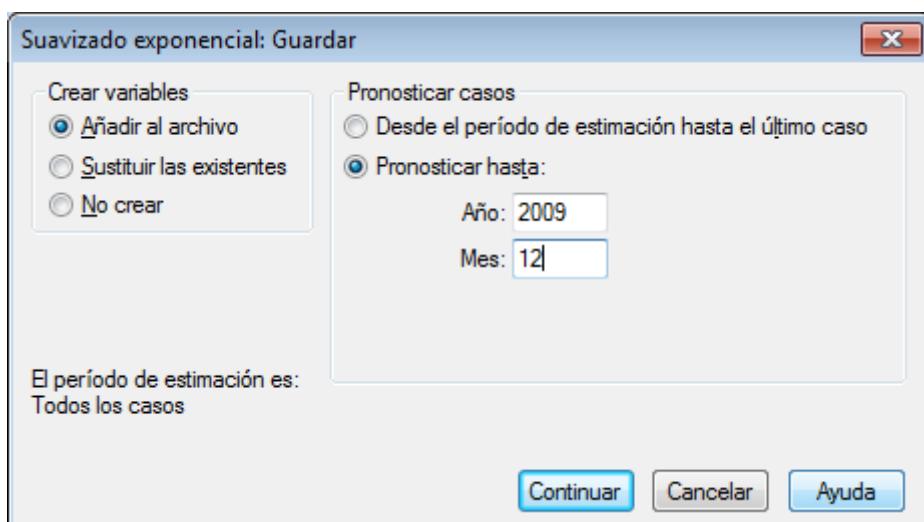
Pincharemos en analizar -> Series temporales -> Gráficos de secuencias y obtendremos las siguientes graficas:



Como se observa los residuos que nos devuelve la técnica de alisado exponencial son similares (en cuanto a magnitud global) a los residuos que nos devuelven los modelos cuadrático y cúbico del análisis clásico. Sin embargo, en el caso de la técnica de alisado los residuos parecen distribuirse de forma más aleatoria que en el caso del análisis clásico.

11. Determina las predicciones de la ocupación hotelera para el próximo año 2009. (Indica sólo las predicciones para los meses de enero, agosto y diciembre).

Usaremos el menú analizar -> Series temporales -> Suavizado exponencial y marcaremos las siguientes opciones para pronosticar los meses deseados.



DATE	FIT_3
JAN 2009	71293,17803
FEB 2009	91189,86559
MAR 2009	109514,40261
APR 2009	114930,28055
MAY 2009	116708,15286
JUN 2009	117235,87414
JUL 2009	125768,87113
AUG 2009	131468,48152
SEP 2009	125406,14000
OCT 2009	112534,07130
NOV 2009	99786,10729
DEC 2009	81166,26742

12. Si queremos que nuestro método de alisado de más importancia al cambio de tendencia que se produce al final de la serie, ¿cómo deberíamos seleccionar los parámetros de alisado?. Justifica tu respuesta.

Deberíamos haber elegido un valor del parámetro γ distinto de 0 ya que el cambio de tendencia al final del modelo nos indica que los últimos valores de este deberían haber tenido un mayor peso en el cálculo final de las predicciones de nuestro modelo. Para este caso podríamos seleccionar el modelo 5, con parámetros $\alpha=0.3$ $\gamma=0.1$ y $\delta=0.0$